# Journal of Educational Psychology

RAYMOND G. KUHLEN, EDITOR
*Syracuse University*

Advisory Editors

JOHN B. CARROLL
*Harvard University*

LEE J. CRONBACH
*University of Illinois*

FRANCIS J. DIVESTA
*Syracuse University*

N. L. GAGE
*Stanford University*

CHESTER W. HARRIS
*University of Wisconsin*

ERNEST R. HILGARD
*Stanford University*

NICHOLAS HOBBS
*George Peabody College for Teachers*

DOROTHEA MCCARTHY
*Fordham University*

CARSON MCGUIRE
*University of Texas*

PAUL H. MUSSEN
*University of California, Berkeley*

T. ERNEST NEWLAND
*University of Illinois*

H. H. REMMERS
*Purdue University*

JULIAN C. STANLEY
*University of Wisconsin*

J. B. STROUD
*State University of Iowa*

DONALD E. SUPER
*Teachers College, Columbia University*

GEORGE G. THOMPSON
*Ohio State University*

BRENTON J. UNDERWOOD
*Northwestern University*

## VOLUME 54, 1963

HELEN ORR
Managing Editor

JOYCE J. CARTER
Editorial Assistant

ELIZABETH S. REED
Advertising Manager

PUBLISHED BIMONTHLY BY:
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 SIXTEENTH STREET, N. W.
WASHINGTON, D. C. 20036

Copyright, 1963, by the American Psychological Association, Inc.

# CONTENTS OF VOLUME 54

Alwitt, Linda F. Decay of Immediate Memory for Visually Presented Digits among Nonreaders and Readers ... 144

Anderson, Richard C., and Anderson, Richard M. Transfer of Originality Training ... 300

Anderson, Richard M. See Anderson, Richard C.

Astin, Alexander W. Differential College Effects on the Motivation of Talented Students to Obtain the PhD ... 63

Astin, Alexander W. Further Validation of the Environmental Assessment Technique ... 217

Ausubel, David P., and Youssef, Mohamed. Role of Discriminability in Meaningful Parallel Learning ... 331

Ausubel, David P. See Fitzgerald, Donald.

Baker, Bruce L. See Levin, Gerald R.

Bassler, Otto C. See Gagné, Robert M.

Berkowitz, Howard. See Siegel, Laurence.

Blocker, Clyde E. See Richardson, Richard C., Jr.

Braley, Loy S. Strategy Selection and Negative Instances in Concept Learning ... 154

Campbell, Vincent N. Bypassing as a Way of Adapting Self-Instruction Programs to Individual Differences ... 337

Capretta, Patrick J., Jones, Reginald L., Siegel, Laurence, and Siegel, Lila C. Some Noncognitive Characteristics of Honors Program Candidates ... 268

Capretta, Patrick J. See Siegel, Laurence.

Cartledge, Connie J., and Krauser, Edwin L. Training First-Grade Children in Creative Thinking under Quantitative and Qualitative Motivation ... 295

Cattell, Raymond B. Theory of Fluid and Crystallized Intelligence: A Critical Experiment ... 1

Crites, John O. Vocational Interest in Relation to Vocational Motivation ... 277

Curtis, Elizabeth Ann. See Siegel, Alberta Engvall.

Finger, John A., and Schlesser, George E. Academic Performance of Public and Private School Students ... 118

Fitzgerald, Donald, and Ausubel, David P. Cognitive versus Affective Factors in the Learning and Retention of Controversial Material ... 73

Fredricks, Robert C. See Otto, Wayne.

Furth, Hans G. Conceptual Discovery and Control on a Pictorial Part-Whole Task as a Function of Age, Intelligence, and Language ... 191

Gagné, Robert M., and Bassler, Otto C. Study of Retention of Some Topics of Elementary Nonmetric Geometry ... 123

Gnagey, William J. See Ofchus, Leon T.

Green, Donald Ross. See Wilder, Nancy Echolds.

Guthrie, George M. Structure of Abilities in a Non-Western Culture ... 94

Guthrie, George M., and McKendry, Margaret S. Interest Patterns of Peace Corps Volunteers in a Teaching Project ... 261

Harris, Chester W. See Klausmeier, Herbert J.

Honzik, Marjorie P. A Sex Difference in the Age of Onset of the Parent-Child Resemblance in Intelligence ... 231

Husek, T. R. See Sorenson, A. Garth.

Isaacson, Robert L., McKeachie, Wilbert J., and Milholland, John E. Correlation of Teacher Personality Variables and Student Ratings ... 110

Iscoe, Ira. See Semler, Ira J.

Jones, Reginald L. See Capretta, Patrick J.

Jones, Reginald L. See Siegel, Laurence.

KATES, SOLIS L. See YUDIN, LEE.
KLAUSMEIER, HERBERT J. Effects of Accelerating Bright Older Elementary Pupils: A Follow Up ........ 165
KLAUSMEIER, HERBERT J., WIERSMA, WILLIAM, AND HARRIS, CHESTER W. Efficiency of Initial Learning and Transfer by Individuals, Pairs, and Quads ........ 160
KRAUSER, EDWIN L. See CARTLEDGE, CONNIE J.
KRAUSKOPF, C. J. Use of Written Responses in the Stimulated Recall Method ........ 172
LEHMANN, IRVING J. Changes in Critical Thinking, Attitudes, and Values from Freshman to Senior Years ........ 305
LEVIN, GERALD R., AND BAKER, BRUCE L. Item Scrambling in a Self-Instructional Program ........ 138
LOCKE, EDWIN A. Some Correlates of Classroom and Out-of-Class Achievement in Gifted Science Students ........ 238
MCKEACHIE, WILBERT J. See ISAACSON, ROBERT L.
MCKENDRY, MARGARET S. See GUTHRIE, GEORGE M.
MARSHALL, HELEN R. Differences in Parent and Child Reports of the Child's Experience in the Use of Money ........ 132
MILHOLLAND, JOHN E. See ISAACSON, WILBERT J.
MUEHL, SIEGMAR. Relation between Word-Recognition Errors and Hand-Eye Preference in Preschool Children ........ 316
OFCHUS, LEON T., AND GNAGEY, WILLIAM J. Factors Related to the Shift of Professional Attitudes of Students in Teacher Education ........ 149
OLSON, LEROY A. Concept Attainment of High School Sophomores ........ 213
OTTO, WAYNE, AND FREDRICKS, ROBERT C. Relationship of Reactive Inhibition to Reading Skill Attainment ........ 227
PECK, ROBERT F. See VELDMAN, DONALD J.
REID, IAN E. See WALLEN, NORMAN E.
RICHARDSON, RICHARD C., JR., AND BLOCKER, CLYDE E. Note on the Application of Factor Analysis to the Study of Faculty Morale ........ 208
SCHLESSER, GEORGE E. See FINGER, JOHN A.
SCHOER, LOWELL. Effect of Intralist Item Similarity on Paired-Associate Learning by Learners of High and Low Verbal Ability ........ 249
SECHREST, LEE. Implicit Reinforcement of Responses ........ 197
SEMLER, IRA J., AND ISCOE, IRA. Comparative and Developmental Study of the Learning Abilities of Negro and White Children under Four Conditions ........ 38
SIEGEL, ALBERTA ENGVALL, AND CURTIS, ELIZABETH ANN. Familial Correlates of Orientation toward Future Employment among College Women ........ 33
SIEGEL, LAURENCE, SIEGEL, LILA CORKLAND, CAPRETTA, PATRICK J., JONES, REGINALD L., AND BERKOWITZ, HOWARD. Student's Thoughts during Class: A Criterion for Educational Research ........ 45
SIEGEL, LAURENCE. See CAPRETTA, PATRICK J.
SIEGEL, LILA C. See CAPRETTA, PATRICK J.
SIEGEL, LILA CORKLAND. See SIEGEL, LAURENCE.
SORENSON, A. GARTH, HUSEK, T. R., AND YU, CONSTANCE. Divergent Concepts of Teacher Role: An Approach to the Measurement of Teacher Effectiveness ........ 287
STRICKER, LAWRENCE J. Note on Social Desirability Response Style and Learning ........ 52
STROM, ROBERT D. Comparison of Adolescent and Adult Behavioral Norm Properties ........ 322
TEAHAN, JOHN E. Parental Attitudes and College Success ........ 104
TRAVERS, ROBERT M. W. See VAN WAGENEN, R. KEITH.
TRAVERS, ROBERT M. W. See WALLEN, NORMAN E.
VAN WAGENEN, R. KEITH, AND TRAVERS, ROBERT M. W. Learning under Conditions of Direct and Vicarious Reinforcement ........ 347

Veldman, Donald J., and Peck, Robert F. Student-Teacher Characteristics from the Pupils' Viewpoint ........ 346
Wallach, Michael A. Perceptual Recognition of Approximations to English in Relation to Spelling Achievement ........ 57
Wallen, Norman E., Travers, Robert M. W., Reid, Ian E., and Wodtke, Kenneth H. Relationships between Teacher Needs and Teacher Behavior in the Classroom ........ 23
Wiersma, William. See Klausmeier, Herbert J.
Wilder, Nancy Echols, and Green, Donald Ross. Expression of Concepts through Writing and Drawing and Effects of Shifting Medium ........ 202
Williams, Joanna P. Comparison of Several Response Modes in a Review Program. 253
Wittrock, M. C. Effect of Certain Sets upon Complex Verbal Learning ........ 85
Wittrock, M. C. Response Mode in the Programing of Kinetic Molecular Theory Concepts ........ 89
Wittrock, M. C. Verbal Stimuli in Concept Formation: Learning by Discovery .... 183
Wodtke, Kenneth H. See Wallen, Norman E.
Youssef, Mohamed. See Ausubel, David P.
Yu, Constance. See Sorenson, A. Garth.
Yudin, Lee, and Kates, Solis L. Concept Attainment and Adolescent Development. 177

*Journal of Educational Psychology*
1963, Vol. 54, No. 1, 1–22

# THEORY OF FLUID AND CRYSTALLIZED INTELLIGENCE:

## A CRITICAL EXPERIMENT[1]

RAYMOND B. CATTELL
*University of Illinois*

The theory of fluid and crystallized intelligence propounded 20 years ago by the author has since accumulated support. However, the crucial issue of whether 1 or 2 general factors subtend intellectual performances has lacked an experiment adequately designed for accurate, determinate, simple-structure rotation at the 2nd order. By factoring culturally embedded with culture-fair intelligence measures on a background of pure personality primaries (N = 277 7th and 8th grade boys and girls), it is shown that 2 general factors indeed exist. A review, with some mathematical formulations, is given of the theory's implications for the nature-nurture ratio, brain injury, standard deviation of the IQ, growth curves, the concept of a relational difficulty hierarchy, test standardization, and the relative validities of traditional and culture-fair intelligence tests.

### PRESENT TRENDS IN INTELLIGENCE CONCEPTS

Since Spearman's demonstration of the general ability factor (g) first put intelligence testing in the framework of scientific research, the field has experienced three major theoretical developments, which, however, have not hitherto been interrelated.

1. Thurstone's (1938) proof that any comprehensive array of ability measures will factor uniquely into a set of simple-structure, oblique, primary abilities, from which g can be obtained as a second-order factor, more uniquely determinable than by the first-order approach.

2. The notion, most frequently associated with the name of Piaget (1947) but widely surmised, and stated very clearly by Ferguson (1956), for example (see also the review by Hunt, 1961), that the formal structure of abilities is partly developmentally determined by what might be called "generalized solution instruments" (i.e., habits the acquisition of which becomes a key, opening the way to rapid advance in some general area of cognitive problem solving). These we shall briefly call "aids."

3. The theory of "fluid and crystallized general abilities," which states that with more refined analytical methods the general ability factor now measured by intelligence tests will be found to be not one factor but *two*. In spite of marked cooperativeness of loading pattern (Cattell, 1952), which makes them difficult to separate, they have properties differing in vital ways for educational and clinical prediction.

This article puts forward entirely new evidence specifically on the theory of fluid and crystallized g factors, but integrates all three of the above and has implications for practical testing in terms of culture-fair intelligence tests.

---

[1] The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare. We are also indebted to D. W. Dunnan, and the school teachers of Springfield, Illinois, and to C. C. Newman, and the school teachers of Paxton, Illinois.

## THEORY OF FLUID AND CRYSTALLIZED GENERAL ABILITY

Although the theory of fluid and crystallized general ability was first stated 20 years ago (Cattell, 1941, 1943), and since developed more precisely (Cattell, 1950, pp. 477–491; Cattell, 1957a, pp. 871–879; Cattell, 1957b), the momentum of custom among intelligence testers, and the absence of sufficiently sustained research on the theory, have hitherto denied to practical test construction and use the advantages in understanding that might be expected to acrue from such dual measurement. The main problem in *research* advance is one of organization, since, as the following summary of the theory will show, coordination is necessary among specialists and practitioners *in several different fields*. Parts of the following theory have been stressed by certain investigators, e.g., in Ferguson's (1956) use of "ability" and "learning set," Hebb's (1942) notion of A and B components in ability, Newland's[2] (1962) use of process and product, and Hayes' (1962) notion of "experience-producing-drives" (though the last joins cognitive and dynamic components here kept separate). But the necessary experimental conditions for constructive conclusions are possible only if the total theory is kept in focus, as is happily the case in, for example, the recent work of Horn (unpublished). The interrelated aspects of the theory may be briefly stated as follows:

1. That among the primary abilities in the area semantically designated "intelligence" in our culture (and which, as Guilford points out, is not *totally* general to cognitive performance) there is not *one* "general ability" second-order factor, as in the Spearman-Thurstone resolution (Cattell, 1943), but more. Two of these are highly cooperative (Cattell, 1952) in the sense that they agree in loading positively most of the general ability primaries and have largely zero loadings outside the intelligence field. Being cooperative they are very difficult to separate, and most previous ability factorings have been satisfied to recognize the joint factor as *one* factor. This conglomerate has been made the basis of IQ measurement and has been called "general intelligence." Technical reasons for this confusion are discussed in the section "Implications for Experimental Checks on the Theory."

2. The two factors, which will be called *fluid* (f) and *crystallized* (c) general ability, $g_f$ and $g_c$, and indexed[3] as U.I. (T) 1 and U.I. (T) 2, have different properties, as follows:

   a. Crystallized ability loads more highly those cognitive performances in which skilled judgment habits have become crystallized (whence its name) as the result of earlier learning application of some prior, more fundamental general ability to these fields. Thurstone's Verbal and Numerical primaries, or achievement in geography or history, would be examples of such products. Fluid general ability, on the other hand, shows more in tests requiring adaptation to new situations, where crystallized skills are of no particular advantage.

---

[2] In communicating to nonpsychologist teachers, for example, the particular aspect of the fluid and crystallized ability theory expressed in Newland's "process" and "product" proves to be readily conveyed by these terms. In any more precise and comprehensive context the terms mislead, however, for the *capacity to learn* is not only a function of the process (fluid ability) but often of the product, as well as of personality and motivation. Particularly when aids have been acquired, the rate of learning is a function not only of fluid ability but of the crystallized products of former application of process.

[3] In the Universal Index (U.I.) system proposed elsewhere (Cattell, 1957c; Cattell & Warburton, 1963), to preserve factor identification during changing explanatory development, U.I. (T) 1 was assigned to general ability and U.I. (T) 2 to flexibility of closure, in French's (1951) series. However, as pointed out elsewhere (Cattell, 1957a, p. 830), there is increasing evidence that flexibility of closure is really the Personality Factor U.I. (T) 19, outcropping in ability when ability tests only are used. Accordingly it preserves the historical importance of *order* in the index much better if the U.I. (T) 2 niche—flexibility going to U.I. (T) 19—is now allocated to crystallized general ability, and flexibility removed to U.I. 19.

b. Before biological maturity (15–20 years of age) individual differences in the difference of $g_f$ and $g_c$ will reflect mainly differences in cultural opportunity and interest. Among adults these discrepancies will reflect also differences in age, since the gap between $g_c$ and $g_f$ will tend to increase with experience and the time decay of $g_f$. Although the speed of an intellectual performance is not a function of U.I. 1 and U.I. 2 alone, (for example, U.I. 22, Cortertia, is a powerful personality contributor; Cattell, 1957a), performances at speed, in any cognitive field, will load $g_f$ more than $g_c$, while tests *without* time limit (unless they are absolutely new and adaptive), such as Thorndike called *power* tests, will have relatively large loadings of $g_c$.

c. Fluid and crystallized abilities will differ on their age trend plots for the general population, $g_f$ reaching an early maximum at 14–15 years, while $g_c$ increases to 18, to 28, or beyond, depending on the cultural learning period for the given subculture. Thereafter, whereas $g_f$ will decline from about age 22 continuously to old age, $g_c$ will show a later and lesser drop. In all these comments on $g_c$, it must be understood that we are talking about the *general* factor in a broad array of crystallized abilities, not any single crystallized ability.

d. The standard deviation of the IQ calculated "classically" from mental ages, will be about 12–16 points for $g_c$ and 24–25 points for $g_f$ (Cattell, 1937, 1955, 1957b). The latter will be more constant and biologically determined, whereas the former will hinge on culture habits. For example, classroom systems which slow the bright and press the dull will reduce the possible absolute individual difference in $g_c$ at any age level, and therefore, $g_c$ standard deviation of IQ, when annual increment is constant. Culture may similarly influence the sigma on $g_c$ in that a powerful education, with larger annual mean increments and, therefore, larger age performance sigma, will tend to reduce IQ sigma. Parenthetically, since traditional intelligence tests are a mixture of $g_c$ and $g_f$, the standard deviation of the IQ on such tests will be a function as follows:

$$^{\sigma}\mathrm{IQ}_c = k\sqrt{\frac{\sigma^2_{f(p)} + \sigma^2_{c(p)} + 2r_{cf(p)}\sigma_{c(p)}\sigma_{f(p)}}{\sigma^2_{f(a)} + \sigma^2_{c(a)} + 2r_{cf(a)}\sigma_{c(a)}\sigma_{f(a)}}} \quad [1]$$

where p subscripts mean across people at one age, and a subscripts mean across age (years), taken as the variance of the means when 1 year samples are included across an agreed number of yearly age groups. As $\sigma_{f(p)}$ and $\sigma_{f(a)}$ can be assumed to be biologically given, the IQ sigma on a traditional, culture-contaminated intelligence test will in general become smaller (over school years only) where schools produce a large age increment in scholarship and where opportunity is not highly adapted to native ability differences, i.e., when $r_{cf(p)}$ is small.

e. For any same-age group the nature-nurture variance ratio will be much higher for $g_f$ than $g_c$ on the hypothesis that $g_f$ is directly physiologically determined whereas $g_c$ is a product of environmentally varying, experientially determined investments of $g_f$. At present the only available, and still scant evidence, reached with the new multiple abstract variance analysis method (Cattell, Stice, & Kristy, 1957) for culture-fair tests (which we shall show are largely $g_f$ measures), does point to higher inheritance[4] for $g_f$. However, although it is our hypothesis that $g_f$ is biologically and physiologically determined, as a function of total cortical cell count, this does *not* mean that one would expect anything like complete hereditary determination. For environment includes gestation period influences and later physical trauma and physiological change, all affecting $g_f$.

f. The reversible, day-to-day or month-to-month fluctuations in level will be different in $g_f$ and $g_c$, the former varying only with general physiological efficiency (Pribram, 1960) and the latter less with physiology and more with recent exercise and interest.

g. The effect of general brain damage (e.g., arteriosclerotic) will be similar in both but more pronounced in $g_f$. The effect of *local* brain damage will be always to lower $g_f$ in proportion to the magnitude of the damage, but the effect on the $g_c$ score will only be through the *localized* ability changes, e.g., verbal aphasia (Lansdell, 1962). Consequently, if the $g_c$ level happens to be estimated from a set of primaries which do not involve the particular area, as a representative, in the standard array of sampled areas in a traditional intelligence test, no significant fall in $g_c$ IQ will necessarily be detected following local injury.

h. The implication in Part a above, that crystallized ability will be a general factor, i.e., covering a substantial total and variety of ability measures, needs to be examined. For such a factor to arise it is necessary that people (*a*) compete in, i.e., be interested and active over, the same series of experiential areas; and (*b*) experience difference degrees of learning opportunity (to acquire skills) *simultaneously* over all these areas. Evidence for the capacity of common instruction and experience to generate such group or general factors comes from the pioneer researches of Anastasi (1936) and of Woodrow (1938); and later work by Greene (1953), Fleishman (1954), and others. Only when the common variance of experience is considerable would we expect the magnitude (mean variance contribution over a standard sample of variables) of $g_c$ factor to exceed that of $g_f$.

Since the present $g_f$ levels of a set of people are simple growth-plus-accident functions of their $g_f$ levels obtaining a few years earlier, and since their present $g_c$ levels are only a somewhat more complex function of $g_f$ levels over preceding years (see Equation 3), one would theoretically expect $g_f$ and $g_c$ to be substantially correlated. The conditions favorable to most substantial correlation prevail in subjects at school age with respect to $g_c$ factor measured over school subjects. In adult life, though school subjects will continue to correlate by "remnant correlations" (see below), a more valid measure of $g_c$ will be made by weighting a much wider range of performances, though any one is likely to be less loaded than were school subjects. The use of culture-fair tests, to measure $g_f$, becomes, under these adult conditions especially, the more reliable, valid, and intelligible intelligence measure.

---

[4] This conclusion requires some special technical discussion for which we have little space. Previous studies, such as those of Blewett (1954b) and Burks (1928), Freeman, Eysenck, and others used *twin* methods in which the environmental difference range (that between twins) is smaller than in the total family of sibs, as in our studies. Since the actual ratios we obtained with Culture Fair Tests are close to the mean of those they obtained with traditional tests, despite our environmental range being larger, the Culture Fair Tests can be concluded to have lower environmental determination. Our ratios were $\frac{122.5}{175.4} = 70\%$ genetic contribution for twins, and $\frac{83.3}{90.0} = 90\%$ for the genetic component in *between-family* variance in $g_f$ (Cattell et al., 1957, p. 154).

i. The rate of learning in any particular area will be a function, as far as cognitive components only are concerned, not only of $g_f$, but also of $g_c$, and the level of the specific ability factor in that area. In so far as $g_c$ and the specific are concerned, we are thus saying that learning begets learning capacity. Even so, the effect of $g_c$ as a general factor should prove to be small relative to $g_f$ in proportion as the learning is in a completely new area. And in old areas the non-$g_f$-predicted variance is more likely to depend upon crystallized group factors than the general factor, namely, those associated with what we have called aids. The demonstration of such action in learning increments, however, involves complex and refined statistical treatment, as discussed by Humphreys (1959), Lord (1956), Tucker (1959), Woodrow (1939), and others.

## IMPLICATIONS FOR EXPERIMENTAL CHECKS ON THE THEORY

The original theory of fluid and crystallized general ability has just been briefly recapitulated in the form of nine propositions integrating quite diverse fields of psychological observation, and indicating a possible array of tangible experimental research checks. However, there are further implications which cannot be stated concentratedly in the nine propositions and which require freer discussion, before the need for certain features in our own experimental design can be clear.

A first major implication is that the form of fluid ability—as a pattern of expressions and loadings—is due to an influence present and operative at the time of the experiment, whereas crystallized ability has a form determined by, and representing, history. Further, one is biologically, the other culturally determined. Our earlier analogy for $g_c$ (Cattell, 1943) of the shape of a dead coral formation, representing the properties and accidental history of once living coral organisms, and Newland's (1962) recent use of "process" and "product" (or Guttman's "analysis" and "achievement") for $g_f$ and $g_c$, bring out this meaning figuratively, in simple metaphors, but not quite as precisely or completely as the real scientific model needs to do.

In behavioral terms, the theory of $g_f$ is that it is "a capacity to perceive relations and educe correlates" in Spearman's (1932) original sense. Further the theory postulates that relations, beginning with those between sensory elements, can be arranged objectively in a hierarchy[5] of increasing abstraction and complexity, and that the level of $g_f$ of an individual at a given time is defined by the position in the hierarchy at which his insight fails, as shown by the onset of chance response behavior. It is a task for logicians (Braine, 1959), ecologists, and machine-simulation specialists to develop bases for this hierarchy; but our hypothesis would be that it will be found to rest on sensory area fundaments, and therefore, require for its genesis intact sensory neurological structure, as stated in Footnote 5. Meanwhile psychologists can delineate a rough hierarchy by trial-and-error item construction, ordered by enlightened difficulty-level analysis (extremity-vector analysis; Cattell, 1957, p. 360) experiments.

---

[5] The present writer's definition of intelligence proceeds to developments of Spearman's relational theory which should be distinguished from those for which Spearman would be responsible. It assumes that the hierarchical organization of relationships among relationships will begin with sensory fundamentals. The special sense areas will therefore form subsets, in the sense of mathematical set theory, and in a psychological sense, in larger sets. From this one might be inclined to conclude that the most complex relations must necessarily lie in the Cartesian space involving all six sensory sets. This *may* be, but at least for the ordinary ranges of intelligence, sufficient complexity can be reached in the hier-

Most attempts, operationally, to measure intelligence, even when it has been conceived as some sort of constitutional, fluid-ability, relation-eduction capacity, have been made, however, through first measuring crystallized ability (i.e., on fundaments and even relations upon which strong cultural training has already occurred). The new movement which began with studying relation-education in perceptual tests (Cattell, 1931, 1940; El Koussy, 1938; Fortes, 1932; Line, 1931; Penrose & Raven, 1936) apparently supports the theory, however, that quite complex intelligence-demanding relations can be successfully set up in entirely new (and therefore *perceptual*, not memory-stored) material. Both experience and psychological reasoning in this area of endeavor lead to the general conclusion, however, that to place individuals in their proper rank order on $g_f$ it is necessary that the test fundament material should be either, (*a*) equally overlearnt by all, or (*b*) equally new to all (and therefore "perceptual" in the sense used in this particular field). Anything in between will contaminate the variance with experiential variance. The sensory area, e.g., whether an analogies test is visual or by sound, seems unimportant. However, the approach through *a*—equally overlearnt—is risky unless certain conditions are watched, and even though it *ranks* persons correctly, it gives no correct account of absolute level. Moreover, there is always the risk that individuals highly familiar with and trained in a field will know *by rote* the correct answers to relational decisions into which they do not have, and have never had, any insight. For example, many people can respond that Einstein's famous formula is $e=mc^2$, not $e=mc^3$.

With these introductions to the general theoretical position we are ready to bring it to precise formulation, and to ask what the implications are for a crucial factor analytic experiment. At this level of precision it can no longer be assumed, at least in the ordinary circumstances of experimental testing of an intellectual performance, that performance level is going to be determined *only* by fluid and crystallized general ability. The specification equation must put these hypothesized two general abilities in the full context of specific abilities and temperamental and dynamic traits, as follows:

$$p_j = s_{j1}F_1 + s_{j2}F_2 + s_{jn}F_N + s_{jt}T_T + s_{si}T_I + s_jF_J \quad [2]$$

where $F_1$ and $F_2$ are $g_f$ and $g_c$, respectively (as in the U.I. factor index); the Subscripts T, I, and J represent

---

archy within the relations arising from *one* sensory area, to reach the limen of failure of insight. This happens because there is a very large number of further subsets among which complex relations can hold, even within one sensory area. The successful use of Culture Fair Tests (involving largely spatial *fundaments*) but transcending merely spatial relations and ability, offers some proof of a sufficiently high relational complexity being attainable upon a limited (two sense) sensory foundation.

temperament, interest (dynamic, motivation), and local area (specific) ability factors, respectively, and $F_N$ represents a factor of strength of immediate memory (for if, as in senility, the capacity to hold the material is deficient the analytical capacities cannot act upon it).

Actually these latter, in a fully expanded formula, would each be several factors, but that expansion is unnecessary for our present statement of essentials. If, moreover, the performance ($p_j$) is made a highly intellectual one, and there is good, uniform, experimental motivation, and sheer speed is not so important that it brings in temperament factors, the analysis of the variance in $p_j$ can be restricted largely to the three cognitive factors.

To understand better what would be the theoretically expected outcome of a factor analysis of purely cognitive variables (i.e., of "intellectual" performance measured under such conditions) let us, however, pry more closely into the nature of $F_2$. This is our specification equation notation for $g_c$ which, according to Propositions e and i, should itself be resolvable *historically* into other factors. In fact, any variable likely to be highly loaded on $g_c$ might, by special factor analysis directed to splitting $F_2$, be expected to yield factors consistent with the Equation 3. This equation states that a crystallized ability performance ($p_{jc}$) is a function of previous time (T) and interest (I) invested in the application of fluid ability; memory; and specific, problem solving aids.

$$p_{jc} = w_{jT}(T_e + T_{ej})\, w_{jI}(I_e + I_{ej}) [s_{j1}F_{1e} + s_{jM}F_{Me} + s_{jA}F_{Ae}] + C_j \quad [3]$$

where:

$T_e$ represents the time given earlier (e) to applying one's abilities in a *whole* field of intellectual performances—i.e., to $p_{jc}$, $p_k$, and the numerous other performances covered by $F_2$;

$T_{ej}$ is time specifically given only to $p_{jc}$;

$I_e$ represents the mean strength of interest in learning at that earlier time which, like T, is assumed common to several $p_{jc}$'s—e.g., those in school work;

$I_{ej}$ is interest specific to $p_{jc}$ only;

$F_{1e}$ is the level of fluid general ability over this *earlier* (not the present) learning period;

$F_{Me}$ is a memory capacity factor, also at its earlier level at that time;

$F_{Ae}$ is an aid, i.e., a discovered "instrument for problem solving" mentioned earlier[6] which operates as a group factor over some area less wide than the general ability factors;

$C_j$ is a statement of the change, an increment or deterioration by

---

[6] An aid is a term for the generalizable habit, discussed by Piaget (1947), Hunt (1961), Hayes (1962), Forgus (1955), Ferguson (1956), and others which opens up faster learning in a whole region, e.g., the infant's grasping the possibilities of babble, the kindergartener's discovery that doing to others as one would like to have done to oneself is an avenue to social goodwill, the junior high school student's realization that a triangle drawn on the board means not that particular triangle but a class of triangles, or that letters in algebra can stand for numbers, and so on. Such a piece of mental equipment stands in the class of achievement-attainment measures, as the use of $F_A$ is intended to remind us. In the last resort, it is itself a product of $F_1$, $F_M$, etc., and belongs to that subclass of attainments which have the special property of aiding further attainment. As such it has an independence of these other terms and a characteristic field of operation, probably as a group factor like a Thurstone Primary Ability. In view of the length of time which Piaget and his followers have been discussing such action it is surprising that there are still no multivariate experimental studies demonstrating their existence, and their nature and position among abilities generally; but they seem a reasonable speculation.

"forgetting" since the original learning experience, due to time and changing interests, retroactive inhibitions, etc. (In the scholarship area one may think of the effect upon a foreign language area of neglect for a decade.)

The s's are the usual situational indices showing by weights specific to the time, interest, ability, etc., how much intelligence, memory, etc., contribute to performance, $p_{jc}$. The w's are weights similarly given to number of repetitions (time of exercise) and strength of interest, e.g., $w_{JI}$ is the rate at which increased time expenditure brings increase in $p_{jc}$ score.

Two statements in this formula need especial comment: (*a*) Whereas in a present performance the specification equation for the factor analytic model *adds* factors of ability and interest to estimate the performance, we have here multiplied ability, memory, etc., by interest and by the time, recognizing thereby a different quality in them from additive influences within the organism. Neither the rather hollow science of learning theory (still helpless where human higher abilities are concerned) nor the degree of precision of factor analysis (Tucker, 1959), permit decision yet as to whether an additive or product model is better. But the product formulation is consistent with our formulations of motivation in the *dynamic calculus*, elsewhere (Cattell, 1957a) and is contingently adopted for these *accumulative learning effects*. (*b*) The T and I terms as indicated by their appearance outside the bracket, are not given separate values for the $g_c$, memory factor and aid factors, though in an expanded formulation they might be given different weights. By the nature of the learning situation intelligence, memory, etc., would normally operate together so they are gathered within the brackets with T and I outside. (The TI, if one wishes, is an integration of interest over time.) The introduction of the memory term ($F_M$) recognizes, as pointed out earlier, that a subject's correct judgment, even on a test item of a highly intellectual kind, may depend on goodness of rote memory, operating over the course of exposure to an intellectual education. For example, a correct decision in geometry may depend on remembering the formula for the area of a circle, or solving a crossword puzzle on good rote memory for Greek. As Cohen (1957) shows, the variance in intelligence performances from this memory component becomes large in older adults.

Now our problem is to ask to what kind of a factor structure the theoretical, psychological formulation in Equation 3 would be expected to lead. Should we expect distinct factors for T and I? The answer seems to be that although very specially planned and exceptionally executed factor analyses might express a product relation like $T_eI_eF_{1e}$ as an addition of three factors, the normal precision and choice of variables would yield a single general factor loading most highly the variables which have simultaneously shared high interest, much time and much involvement of fluid ability. However, because they are multiplied by the same $T_eI_e$ the intelligence, memory, and aid factors would tend to be cooperative (Cattell, 1952) (i.e., loading the same variables). And if the variance in $T_eI_e$ were great, we should expect a powerful second-order factor among them defined by $T_eI_e$. Thus there are really two possible definitions of the crystallized general ability factor: (*a*) As the first-order factor $T_eI_eF_{1e}$, representing the past areas of investment of fluid ability, and (*b*) as the second-order factor—strictly a general achievement factor—corresponding to the investment of $T_eI_e$ in fluid ability expression, memory, and various aids and special primary abilities.

The situation is further complicated by the fact that skills not only increase but decay, and a common history of change or decay *after* the pattern has been reached as in the first part of Equation 3 could also lead to a common factor structure. This, with some overcondensation, we have represented by the term $C_j$ in the latter part of Equation 3. For example, if $C_j$ is a time function and our sample is scattered from 30 to 60 years of age, a general factor would tend to appear over all school subjects due to their all being forgotten to the same degree, as a function of the time lapse. Such considerations have been grossly overlooked in traditional adult intelligence tests, as well as the fact that a new general crystallized ability factor, distinct from the scholastic one, is likely to arise through common learning by the application $T_eI_e$ to whatever common areas of $g_f$ use exist for adults in our culture. Here only culture-fair, but not traditional, intelligence tests seem to recognize that in the majority of our population which proceeds to "nonverbal" occupations (e.g., garage mechanics, farmers, etc.) a substantial decay of the attained crystallized ability levels in vocabulary and similar areas supervenes (Burt, 1955). Further, the normal adult intense investment of skills in relatively narrow occupations will tend to make the appearance of a second "adult" crystallized ability factor, superseding the school curriculum crystallized general ability factor, a relatively poor substitute, of lower variance and poorer predictive power.

In summary, one must face the complication that the crystallized general ability factor is for various reasons Protean. Strictly the term crystallized intelligence (Symbol $g_c$) should be restricted to the first of the two factors mentioned three paragraphs above, namely, to the Expression $T_eI_eF_{Ie}$. The second-order factor covering $T_eI_e[F_{Ie}+F_{Me}+F_{Ae}]$ is in any sense of the word a *general achievement* factor, though it must be confessed that a substantial part of traditional intelligence test scores derive from this source. Both of these factors are at the mercy of culture, and even before one encounters the difficulty of cross-cultural comparisons thereon, one gets entangled in the shift which occurs in them between the schooling age and the adult activity age. The Protean character is thus at best reducible to four concepts, corresponding to four crystallized general ability factor patterns.

1. School age crystallized intelligence ($g_{cs}$)—common to $g_f$ expressions powered by time and interest within the culturally set framework of the school curriculum.

2. School age crystallized achievement ($a_{cs}$)—a second-order factor across intelligence, memory, special ability areas, etc., corresponding purely to the effect of common time, interest, memory, and curriculum. It will differ markedly from $g_{cs}$ in being predictive of achievements (e.g., athletics, foreign languages, and rote areas of school learning) in which $g_f$ has never had any appreciable importance or role.

3. Adult activity crystallized intelligence ($g_{ca}$)—representing whatever skills develop strictly from fluid intelligence being applied in whatever activity areas receive common time and interest from adults, i.e., $T_aI_aF_{Ia}$.

4. Adult activity crystallized achievement ($a_{ca}$)—representing whatever general factor is produced in adult achievements through their common expenditures of time and interest. Except in mixed samples from different cultures

this would be a very tenuous factor of little practical predictive use.

Only in the case of Concept 1 would we expect the correlation of fluid and crystallized intelligence factors to be really substantial, and the term intelligence is best eschewed completely for Concepts 2 and 4. Also the terms "school practised" and "non-school-practised" abilities are misleading for $g_c$ and $g_f$, for the first would describe both Concepts 1 and 2 and the second could be either Concepts 3 or 4, or $g_f$. Any combination from decay influences is omitted from this reduction to four general factor concepts.

## DESIGN OF THE EXPERIMENT IN RELATION TO EXPECTATIONS

Testing the theory of fluid and crystallized intelligence requires not one experiment but a comprehensive organization of experiments. Yet one which needs to be performed before all others, as crucial to further work, is a factor analysis to check whether one factor or two runs through most intellectual performances.

Technically, two questions need care in planning such a factor analytic experiment:

1. Are we going to aim at abstracting $T_eI_eF_{Ie}$ as a single factor or at splitting it? Theoretically it is not beyond the power of fine factor analysis to split a presently operating whole into the components which historically produced it. A product relation would be approximated by additive factors as the plasmode studied by Bargmann (1955), Cattell and Dickman (1962), Cattell and Sullivan (1962), and Thurstone (1947) sufficiently demonstrate. But in the present case $T_e$, $I_e$, and $F_{Ie}$ would be cooperative, because by the necessary choice of tests for intelligence they would be variables also on which common $T_e$ and $I_e$ have been expended. Besides there is much to suggest that product relations are often revealed by cooperative factor patterns. Consequently, since cooperative factors are hard to rotate correctly, it would seem best at this stage not to attempt such finesse but to accept variables sufficient to reveal crystallized intelligence as a "factor in being," if it so exists. When such deeper analysis is eventually undertaken it would be well to keep in mind that $I_e$ might split again into a number of personality and dynamic factors (e.g., super ego strength, emotional stability) affecting the investment of fluid intelligence in crystallized intelligence skills. Essentially this has recently been powerfully argued by Hayes (1962). Decay factors will also need consideration.[7]

2. Provisions must be made for definition by rotation of *factors possibly general to every cognitive test employed.* One reason why the present theory, rather than the Spearman-Thurstone-Thomson theory of a single general intelligence factor, has not hitherto received due and serious consideration is that most workers in the cognitive field have included only cognitive tests and thus rendered quite impossible any revelation of a distinctly rotatable second

---

[7] A summary of the main influences now posited in any performance in the area semantically designated intelligence might be convenient at this final step: (*a*) a fluid *general* intelligence factor; (*b*) a crystallized general intelligence factor; (*c*) one or more crystallized auxiliaries, corresponding to time and application acting with sheer capacities to memorize at the former among time (e.g., knowledge of vocabulary as distinct from nice judgment of meaning among familiar basic English words); (*d*) factors corresponding to any general patterns of change or decay since learning; (*e*) existing personality and dynamic (motivation) factors helping application to an intellectual test.

Here *b*, *c*, and *d* have structure in virtue of history, while *a* and to some extent *e* are structures due to present organismic functional unities.

general factor. The procedures of Burt (1955) and Vernon (1950) for example, neither permit unique determination of the alleged general intelligence entity nor separation of a possible second general factor.

The technical failure in factor analysis here, by which investigators have failed to see the wood for the trees, is only a special case of ignoring a general principle in scientific method—namely, the figure-ground principle, according to which, if we wish to define X adequately we must introduce data which is not X. In this case *to rotate with any approach to precision a factor expected to be general to all abilities one must include a wide framework of nonability factors (not merely variables) as hyperplane stuff* (Cattell, 1952). Indeed, in spite of the immense importance of the intelligence concept to education, the second-order general ability factors appear never to have been so rotated, in any research discoverable by the writer, to a comprehensively determined simple structure position.[8]

In the present study we planned to reach the second-order analysis with only nine ability measures—a bare but sufficient minimum for testing the main cognitive hypotheses—but with 13 nonability measures. The latter were chosen (*a*) to outnumber abilities, (*b*) to be well-known *factors,* not merely variables, so that they would fan out into a maximum number of dimensions to give stability of rotation of the ability factors with respect to many reference points, and (*c*) themselves to have a known second-order structure, the emergence of which in the present blind rotation would give a check on its general correctness. These requirements[9] we felt to be best met by taking Personality Factors A through $Q_4$ (having the same meaning as on the 16 PF Test) on the High School Personality Questionnaire (HSPQ; Cattell & Beloff, 1959).

The nine ability variables consisted of five Thurstone Primary Abilities—Verbal, Spatial, Reasoning, Number, and Fluency—and four subtests from the Institute of Personality and Ability Testing (IPAT) Culture Fair Intelligence Test, Scale 2a—namely, Perceptual Series, Perceptual Classification, Matrices, and Topology. If our hypothesis is correct all the Thurstone primaries, with the exception of Spatial and possibly Fluency preformances should load on a crystallized ability factor (as also should be the case with each specific aid factor).[10] On the other hand, the Culture

---

[8] In many cases investigators have shown a curious inconsistency in believing that first-order factors should be rotated but second orders should be left where they fall! (The Schmidt-Leiman determination of loadings of variables on second orders, for example, assumes this.) In other cases, the explanation seems to be the usual exhaustion phenomenon, in which, by the time the investigator gets to the second order of a long factor analysis, neither time nor energy remains to press on to a determinate second-order resolution. Mainly, however, it seems that like early climbers on Everest, they have not dragged along at the early stages of the expedition the equipment which foresight might have indicated to be essential at the last stage of ascent.

[9] This procedure of entering the ability research with more personality than ability measures, may seem quixotic to conservative researchers, but in addition to the reasons given above it is actually strategically necessary in terms of demonstrating that second-order ability factors are in fact *distinct* from personality factors. This can by no means be taken for granted, since some "primary abilities," e.g., flexibility of closure, now appear (Cattell, 1957a) as outcroppings of personality factors "in disguise" in the cognitive realm.

[10] A small note on a large issue is necessary here in connection with the notion raised above that an accumulation of separate aid developments might constitute a general factor. The magnitude of the aid factor, $F_A{}^1$ is a joint function of $F_1$ and some environmental "accident" $A^1$, though whether it should be written as (*a*) $F_A = F_1 \times A$, or, (*b*) $F_A = F_1 + A$ is uncertain. In either case A would be operating as

Fair subtests, being purely perceptual and lacking even pictorial reference to the not-present should load only a fluid general ability factor. However, since even the Verbal item decisions involve some immediate adaptive performance, the fluid factor should extend to some degree (in the rotation position defined by the wider realm of personality measures) into such performances as Verbal, Spatial, etc., primaries.

The choice of sample was made on the assumption that Horn's (unpublished) research would use adults. For a complete answer on the above hypothesis requires a check on the expectation that the crystallized intelligence factor will behave very differently (as $g_{cs}$ and $g_{ca}$) with children and adults. There is some advantage in beginning with, say, 12-year-old children, because we may expect less complication from any decay term ($C_j$ in Equation 3). Also one might expect the $g_{cs}$ factor, based on so uniform an area of earlier application, to be larger than with adults (about the same variance as the $g_f$) and thus easier to locate in a first study.

## DATA AND ITS ANALYSIS

The sample consisted of 124 eighth grade boys and girls from Paxton, Illinois, and 154 from the seventh and eighth grades of a junior high school in Springfield, Illinois, a total of 277 (i.e., one dropped) being carried forward to the analysis. The Thurstone Primaries (except for Fluency) were given in both forms, and HSPQ was similarly given in separate A and B scores for each primary personality factor; but the Culture Fair Intelligence Test used only one form (which must be remembered in appraising later saturations). Testing took about 3 hours and was under one examiner.

The *first* order analysis required product-moment correlations[11] for 44

---

a group factor covering from two or three to a large number of performances.

In the early stages of the operation of an aid, as suggested above, it should be detectable in either case as a group factor. In later stages, it would be possible for an accumulation of overlapping group factors to be resolved as a general factor, indeed, *they would constitute part of the crystallized general ability*. That a set of overlapping small group factors could alternatively appear as a general factor has puzzled factor analysts since the early debates of Thomson and Spearman on the interpretation of g as a "sampling of bonds" versus a "single power." This factor analytic problem can be kept clear only by invoking the *principle of organismic setting.*

This principle states that although each of a set of separate covariational entities may appear as a separate group factor, yet, if they all have some quality in common, so that they can appear as part of a larger organization, then, by simple structure resolution, their *sum* will also appear as a single factor, in the context of several other, qualitatively different, such accumulations. Thus although each leaf in a heap of leaves is a separate entity and (if its variation, e.g., in perception from various angles, were considered, a separate factor) yet the heap also has its dimensions, as an entity, among other heaps. Similarly for cells in an organism or people in a team. What one gets factor analytically depends upon the organismic level at which measurements and rotations are made.

Even before one gets a shift of organismic level, a mere change of scale, from microscopic, or, in terms of psychological variables, from low to high density of representation of variables, can convert the resolution from a multiple group factor outcome to a general factor outcome. Pending further experiment our hypothesis is that the variance produced by cummulative action of distinct aids will, except for what is left in the truncated group factors (Cattell, 1962) become part of the crystallized general ability factor, i.e., $F_I$ and $F_A$ in Equation 3 will both go into its variance. This states an important difference from Hunt's (1961) theory.

[11] The correlation matrix, centroid, lambda transformation, and final simple structure matrix have been deposited with the American Documentation Institute. Order Document No. 7381 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

variables (9 Thurstone abilities, 4 Culture Fair subtests, and 28 HSPQ personality variables), i.e., in all 41 plus 3 random variables for hyperplane testing. Since they consisted mainly of paired markers for factors, it was not surprising that the Tucker (Thurstone, 1938) and Guttman (Cattell & Sullivan, 1962) tests (separately applied) pointed to 20–24 factors to be extracted. We settled on 22 as the limit for determinate, converged, iterated communalities and iterated four times to reasonable stability.

The rotation for simple structure was carried out with extreme care, since as will now become apparent, all conclusions in the second order hinge upon the angle values settled upon in the first. Simple structure reached, after four overall rotations, the very high value of 82% of all variables lying within the ±.10 hyperplane. At this point in the blind rotations an inspection showed that the usually accepted structure for both the Thurstone Primary Abilities and HSPQ primary personality factors had emerged, independently, with excellent clarity. The two prescribed markers, and virtually nothing else, loaded each named factor. Although only one marker had been inserted for Fluency and each Culture Fair subtest, these five picked up enough on other variables, e.g., classification had loading on HSPQ Factor B, for each to appear as a nonspecific factor, though they are in the usual wider sense, specifics.

The $C_R$ matrix for this carefully rotated study was now inverted to give the correlations among the 22 factors ($C_F = DC_R^{-1}D$) and this was again factored. The advantage of proceeding to the second order from the $C_R$ instead of from correlations of actual battery measures of the factors is that we eliminate distortions from the unreliabilities and imperfect validities of actual scales. In exchange we pick up whatever uncertainties remain in the simple structure rotation, but with at least 20 variables to define the hyperplane, this should be a lesser risk and offer a better design.

Centroid factoring of the first-order 22-factor matrix yielded, by the same pair of criteria as before, 8 factors. Their simple structure rotation is shown in Table 1, together with $C_F$ of their correlations (see Footnote 11). The simple structure here is no longer quite as clean as at the first order. If the oddity of Factor 7 is omitted (with which we are not much concerned) the count is, however, still 59%, which is high among reported studies and at any rate the essential maximum or plateau attainable among present possible rotations.

The structure of Table 1 reveals unmistakeably the distinctness of the fluid general ability pattern, which is Factor 1 (loading the four subtests of the Culture Fair, the Thurstone Spatial, and the HSPQ Ego Strength, C, factor), and the crystallized general ability factor, Factor 2, which loads Verbal ability, Reasoning, and Number most highly. That these two ability factors, roughly orthogonal to the remaining six, are essentially in the right position is attested further by the correctness of the second-order factors emerging among these latter personality factors. Factor 3 is the typical U.I. Q. I. (Cattell, 1957a) second-order extraversion factor (Carrigan, 1960) covering the primaries A, D(−), F, H, $Q_2$(−). Factor 4 is the general Anxiety versus Adjustment factor—C(−), O, and $Q_4$(−), (with $Q_3$ slightly anomalous); and Factor 5 continues these previously recognized second-order factors to include that indexed as U.I. Q. III (Cattell, 1957a). Although these are to be remarked here only as confirmation of the "setting" for the two main ability factors, it is possible that Factor 6—"super-ego-good-upbringing" or Control—has also some interaction with crystallized ability.

Let us for the moment postpone ex-

TABLE 1

SECOND-ORDER FACTORS AND THEIR INTERCORRELATIONS

Part A: Factor loadings (in reference-vector correlation form)

| First order factor variable | $F_1$ $g_f$ | $F_2$ $g_c$ | $F_3$ Exvia-Invia | $F_4$ Anxiety | $F_5$ U.I. Q. III | $F_6$ Control | $F_7$ ? | $F_8$ ? |
|---|---|---|---|---|---|---|---|---|
| Thurstone primaries | | | | | | | | |
| Verbal | .15 | .46 | −.17 | −.03 | .04 | .07 | −.05 | −12 |
| Spatial | .32 | .14 | −.02 | −.03 | −.03 | .05 | −.27 | .04 |
| Reasoning | .08 | .50 | .08 | .02 | −.02 | .02 | .38 | −.14 |
| Number | .05 | .59 | −.04 | .05 | −.07 | −.05 | −.10 | .03 |
| Fluency | .07 | .09 | −.52 | −.18 | .12 | .26 | .13 | −.07 |
| IPAT Culture Fair subtests | | | | | | | | |
| Series | .35 | .43 | .05 | −.12 | .06 | .10 | .24 | .04 |
| Classification | .63 | −.02 | −.05 | −.05 | .00 | .14 | .16 | .06 |
| Matrices | .50 | .10 | −.23 | .04 | −.01 | −.08 | −.25 | .42 |
| Topology | .51 | .09 | .16 | .12 | .03 | −.03 | −.14 | −.00 |
| IPAT HSPQ | | | | | | | | |
| A Cyclothymia | −.04 | .52 | .31 | −.23 | .05 | .11 | .20 | .03 |
| C Ego Strength | .21 | −.07 | −.04 | −.52 | −.08 | −.09 | −.49 | .12 |
| D Excitability | −.04 | −.44 | −.44 | −.09 | .09 | .04 | .14 | −.27 |
| E Dominance | −.15 | −.01 | −.05 | −.00 | .61 | −.04 | −.02 | −.17 |
| F Surgency | −.05 | .09 | .39 | −.07 | .06 | −.40 | −.20 | −.09 |
| G Super Ego Strength | −.14 | .08 | −.01 | −.10 | −.10 | .43 | .41 | .05 |
| H Parmia | .21 | −.04 | .55 | −.17 | .00 | .13 | .03 | −.03 |
| I Premsia | −.09 | −.29 | −.05 | .26 | −.54 | −.05 | .43 | −.01 |
| J Coasthenia | −.10 | −.04 | .01 | .39 | −.12 | .45 | .01 | .16 |
| Q Timidity | .16 | −.01 | .01 | .73 | .09 | −.04 | .10 | .17 |
| $Q_2$ Self-Sufficient | −.06 | .05 | −.37 | −.05 | −.43 | −.04 | −.04 | .05 |
| $Q_3$ Self-Sentiment Control | .05 | −.02 | .06 | .00 | .32 | .35 | .56 | .11 |
| $Q_4$ High Ergic Tension | −.04 | .37 | .03 | .45 | −.00 | −.37 | −.14 | .05 |
| % in ±10 hyperplane | .51 | .60 | .60 | .51 | .73 | .55 | .32 | .60 |

Part B: Factor correlations

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1. Fluid general ability | 1.00 | .47 | .29 | .35 | .09 | .21 | −.01 | −.15 |
| 2. Crystallized general ability | .47 | 1.00 | .17 | .15 | −.03 | .28 | .03 | −.20 |
| 3. Extroversion versus Invia | .29 | .17 | 1.00 | .17 | −.01 | .10 | −.24 | −.05 |
| 4. Anxiety versus Adjustment | .35 | .15 | .17 | 1.00 | .24 | .06 | .14 | −.12 |
| 5. U.I. Q. III | .09 | −.03 | −.01 | .24 | 1.00 | −.46 | .06 | .26 |
| 6. Control | .21 | .28 | .10 | .06 | −.46 | 1.00 | .07 | −.70 |
| 7. ? | −.01 | .03 | −.24 | .14 | .06 | .07 | 1.00 | −.11 |
| 8. ? | −.15 | −.20 | −.05 | −.12 | .26 | −.70 | −.11 | 1.00 |

perimental examination of our assumptions about the inherent modes of expression of these two factors as shown by loadings and pause to examine their mutual relation as shown in Tables 1 and 2. The substantial positive correlation, +.47, shown in Table 1, Part B, is what would be expected from our theory, on the grounds that both must be highly related to an entity which existed a few years earlier, and which we have symbolized (Equation 3) by $F_{1e}$ (or $g_{fe}$)—the fluid ability level earlier. The present crystallized ability level—$F_2$ (or $g_c$)—is a function of $F_{1e}$ and the time and interest applied at that time.

The present fluid ability level—$F_2$ (or $g_f$)—is an organic continuation of $F_{1e}$, falling short of perfect correlation with it only because different times have elapsed for different people and the interlude has treated them more or less kindly in terms of brain damage, etc.

If next we ask what this correlation implies in terms of a third-order analysis, different theories will give different answers. Our theory would posit that $F_1$ and $F_2$ ($g_c$ and $g_f$) should come together in a single higher order factor, because historically a single influence, $F_{1e}$, is responsible for them both. It has, so to speak, fathered $F_1$ out of time (as "the child is father of the man") by simple maturation, and $F_2$ out of experience by interest and learning investment. Whether the third-order $F_{1e}$ should load higher on $F_1$ or $F_2$ depends on many things. High similarity of experiences (including time length) and interest should raise the loading on $F_2$, whereas heterogeneity of the group in this respect should lower it. Heterogeneity of the group in age and time lapse since the formative period of $g_c$ should lower the loading in $F_1$. In a group of school children such as the present, in the midst of their growth period, the $F_{1e}$ loading in $F_1$ should be quite high and that in $F_2$ a little short thereof. In adults or children more heterogeneous for age, the loadings should both be lower.

The third-order analysis designed to test these hypotheses again had, by reason of the initial planning, enough hyperplane stuff to determine the rotation of this still broader general ability factor. Further, the rotation was again made with pure factor measures, not error-contaminated actual battery scores, for simple structure was attained in the reference vector matrix underlying $C_F$ shown in Table 1. In this $C_F$ four factors were indicated by the Tucker test and a satisfactory 63% in the hyperplane was obtained after five overall rotations, with results shown in Table 2.

TABLE 2

THIRD-ORDER FACTORS AND INTERCORRELATIONS

| Second-order factors | Third-order primary factors 1. | 2. | 3. | 4. |
|---|---|---|---|---|
| Fluid intelligence | .69 | .02 | −.07 | .00 |
| Crystallized intelligence | .63 | −.04 | .07 | .32 |
| Extraversion versus Invia (U.I. Q. I) | .18 | .03 | .38 | .23 |
| Anxiety versus Adjustment (U.I. Q. II) | −.01 | .01 | .00 | .79 |
| U.I. Q. III | .09 | −.51 | −.07 | .32 |
| Control | .01 | .99 | −.03 | −.05 |
| $F_7$ | .06 | .03 | −.74 | .04 |
| $F_8$ | .02 | −.69 | .08 | −.06 |
| % in ±.10 hyperplane | | | | |
| Primary factor correlations at third order | | | | |
| 1. | 1.00 | .38 | .14 | .25 |
| 2. | .38 | 1.00 | .00 | .10 |
| 3. | .14 | .00 | 1.00 | −.16 |
| 4. | .25 | .10 | −.16 | 1.00 |

If our hypotheses is correct it would not be right to conclude from Table 2 that "fluid and crystallized general ability form a single third-order factor," but rather that a single influence, which is fluid ability as it stood during the formative period of crystallized ability, is causative to the present levels of both. That this "formative fluid ability" influence is in fact $F_1$ in Table 2 could be checked by preserved records of the subjects' earlier fluid ability scores. It is noteworthy in Table 2 that this fluid ability factor has no significant loading on any personality factor, but stands as a measure of pure ability. On the other hand, it has slight, significant correlations *as a factor* with control and anxiety of a kind we have come to expect from the linking of intelligence with other factors through the natural selection of social status (Cattell, 1945; Fisher, 1958). The factor contingently dubbed Control here has, on the other hand, significant relations to crystallized ability in Table 1 and must be considered one of the personality factors contributing, partly by aiding school memorizing, to crystallized ability, via $F_M$, etc., in Equation 3 above.

## NATURES OF FLUID AND CRYSTALLIZED INTELLIGENCE

With concepts defined and checked at the experimental level we can now proceed to broader formulations. The statements so far are:

1. That intelligence test and school performances must be considered, more comprehensively than hitherto, to be determined, at the moment, by personality and motivation factors as well as by two distinct second-order general factors, (fluid and crystallized intelligence) as symbolized in Equation 2. Equation 2 is primarily an expression for performance in a timed test, but with trivial adjustments, is also an expression for rate of learning.

2. As a result of this mechanism of learning operating uniformly across school subjects, and other areas, for time intervals differing for different people, third-order general factors of school achievement (for children) and adult activity achievement (for adults) are created. (There is no space to discuss their relation here, but they should be slightly positively correlated.) The term crystallized intelligence, however, is semantically better reserved for the second-order factor in such performances as derive their positive correlations only from the action of fluid ability, not from memory and various special abilities operating in the school (or adult) settings. Crystallized intelligence and achievement are thus *not* identical general factors, the former being narrowed in manifestation to the handling of complex relations, whereas the latter could cover whatever the breadth of the cultural schooling curriculum embraces.

A special logical complexity arises here, however, in that rate of learning in complex fields is assisted both by fluid and crystallized general abilities, as well as by such aids (usually primary, group factor patterns) as are relevant to the particular field (Equation 3). Consequently, crystallized ability must be considered to be begotten by crystallized ability as well as by fluid ability. This may seem to contradict the concept above that crystallized ability is a consequence and function of fluid ability levels, but it is not so if we consider (*a*) that fluid ability is so to speak put out at compound interest, and recognize also that the rate of return may be different over different periods of learning, depending on interest, etc.; and (*b*) that crystallized ability is not *only* a function of fluid ability but also of personality factors, etc. Thus personality factor "deposits" occur in what at the time of learning is considered an ability —crystallized general ability—and also in aids. If space permitted, these relations of earlier personality and motivation factors, as operating in Equation 2 to crystallized abilities as in Equation 3, could be worked out in a fourth or "historical depth" equation.

As to the descriptive similarities and differences of $g_f$ and $g_c$, both are very general to complex, intellectual, relation-perceiving performances, and even in regard to the variables chosen to discriminate them one notices, in Figure 1, that each has some tendency to load the variables set to mark the other. However, the slight loading of crystallized intelligence in say Fluency is not as great as that of fluid ability in Verbal or Series. (Note with oblique axes the lines of projection are drawn parallel to the axes in Figure 1.) Parenthetically, it will be noted that if the experiment had been undertaken without personality factors, etc., as hyperplane substance, the rotator would almost certainly have fallen into the false resolution of making these factors more highly correlated

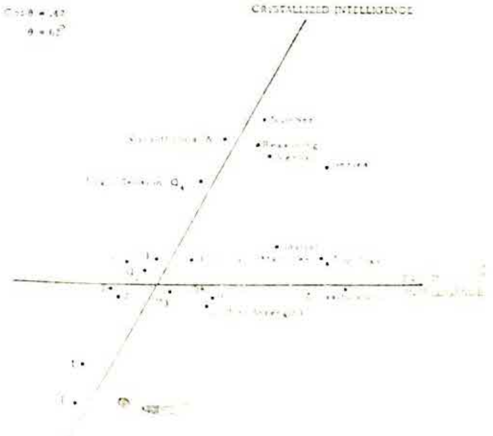


FIG. 1. Culture-fair subtests and Thurstone Primaries in relation to $g_f$ and $g_c$. (Rotation assisted by 16 PF factors in hyperplane.)

(especially in the absence of D, I, A, and $Q_2$) by going through the ability variables only, e.g., through Verbal and Topology.

Figure 1 also confirms the whole argument of Equation 2 to the effect that crystallized intelligence, being partly the product of motivational and personality history, will have more significant association with personality factors. This association is not only statistically significant but makes psychological sense, for crystallized ability correlates substantially positively with Cyclothymia (Factor A in HSPQ), and negatively with Excitability (D), and Premsia (I) which other researches show to be similarly related to general school achievement. Similarly in the motivation realm (Connor, 1961), docility (low self-assertion) and super ego strength have the highest correlation with total school achievement. Their correlations with the Verbal (V) factor, the Reasoning (R) factor, and the Culture Fair are, respectively, .20, .48, and .02; and .19, .41, and −.02, (see Footnote 10) the Culture Fair (fluid ability) measures consistently being more free of association with personality factors affecting achievement. As stated above, this greater freedom of the second-order fluid ability from personality loadings, which supports our hypothesis, is not incompatible with its having correlations therewith at the third order, pointing to genetic and social status ties.[12]

A word is necessary here to forestall a misunderstanding which will almost certainly arise in the attempt of conservatives to avoid any major restructuring of ideas. This will be the view that $g_f$ is nothing more than the k or "practical ability" factor, located by such factorizations as those of Vernon (1950) and Cohen (1957). There are three good reasons for rejecting this. The first, comparatively trivial, is that k has always shown decided associations with masculinity and mechanical experience whereas in our mixed boy and girl sample $g_c$ shows statistically no difference and is actually a shade higher for the girls. Secondly, and more systematically, the whole mode of factor resolution in the Burt-Holzinger-Vernon bipolar system is quite different from multiple-factor simple-structure resolution and never yields a one-to-one matching. The former method puts verbal ability on one side and practical-mechanical on the other. In the present writer's opinion this is in any case a fallacious form of factor resolution, for since the first general factor is indeterminate, the whole hierarchy is fictitious. But, fallacious or not, it is, except by some accidental caricature, too different for $g_f$ ever to be identified with k. Finally, however, $g_f$ is *not* particularly loaded on mechanical aptitude variables (the evidence is not given here, but in Horn, unpublished), or in spatial ability measures, or one-sidedly in any of those performances which have been included in this vague and impractical practical ability artifact.

---

[12] In looking for indications of some separate action of the interest-personality factors posited in Equations 1, 2, etc., one may note in Table 1 that Factors 6 and 7 admit of a single factor being rotated from them which loads the two Super Ego factors, G and $Q_3$; Desurgency, F(−); and positive performance on Reasoning, Fluency, Series, and Classification. It makes good psychological sense that such cognitive performances should be aided by such personality factors, but it suggests they are acting in the present test situation (as in Equation 1) rather than through help in past performance as in Equation 3.

On the last issue—that of giving us a far more complete descriptive differentiation of $g_f$ and $g_c$ in the extent and nature of their loading fields—there is admittedly a great deal still to be found out. Considering the volumes of ability correlating that have gone on over the past 20 years it has been, to say the least, theoretically uninspired. Especially the hypothesis should be tested that $g_c$'s loading pattern will alter with age, subculture, and culture whereas $g_f$'s will not. Already such studies as that of Sisk (1940) point to the effect on the form of the crystallized intelligence pattern of such very local influences as an "interest in study" over particular college courses shared by members of a group. However, one unjustified assumption needs to be scotched which has been, among teachers, as obstructive in holding up the practical introduction of culture-fair tests, as the k illusion has been in buttressing outworn theory. This is the "faith validity" judgment that because such tests employ shapes they must be testing spatial ability. As Figure 1 shows, the loading of the fluid general intelligence factor is not as great in Spatial as in Matrices, Topology, and Series tests, and though such tests depend on perception there is as yet no evidence that they load unduly Thurstone's perceptual factors.

If, as stated at the outset, the theoretical aim of culture-fair intelligence subtest construction is to require complex relation eduction in material that is either completely new to or completely overlearnt by all persons to be tested, it is of interest to discover what these areas may be. A systematic attack on this was begun some years ago in the work of Cattell (1940), Line (1931), Fortes (1931), and others and pointed to *simple* spatial perception, shade and pitch perception, and probably tactile experience and awareness of the body, as being sufficiently overlearnt in most cultures, but this exploration was never continued. Meanwhile, from a practical intelligence test construction standpoint it is interesting to note that the Classification, Topology and somewhat less, the Matrices tests are good measures of fluid ability (at any rate as used in the Culture Fair Test) and Number, Reasoning and Verbal primaries of crystallized ability. Incidentally, the evidence of cultural learning in the Series test is a bit surprising.

While commenting on odd loadings in Table 1 one may note also the loading which a factor, Number 8, outside the general ability factors, has on the Matrices test. The present writer has elsewhere argued, on the theoretical likelihood of such a loading, against the legitimacy of using Matrices alone, for measuring fluid general intelligence, instead of a complete Culture Fair battery averaging several subtests. This shows that indeed appreciable specifics or extra factors are as likely to occur in perceptual as in traditional subtests. It is possible they will not prove to be *cognitive* specifics, but, as the present analysis suggests, less than half of the variance in Matrices appears to lie in general intelligence, and the rest is sheer unknown bias in the test measurement if Matrices alone is used as an intelligence test.

As to the present accuracy of estimate of $F_1$ and $F_2$ the loadings of the subtests in the Culture Fair battery[13] on the fluid ability factor are, in general, as good as, or a little better than, those of the Thurstone primaries on the crystallized ability factor (mean = .40 as against .36).

---

[13] The search for further good subtests might now be guided by the emerging theory that

## NEEDED RESEARCH

At the present moment researches on these theories are as vital in the applied as in the basic field. For if the main concepts are correct, present routine testing practices in school, and still more in vocational selection of adults, are wasting a lot of talent. The traditional intelligence test, because it confounds fluid with crystallized intelligence and even with the third-order general attainment factor, gives a speciously good correlation, at least in schools, with the criterion. A test in June predicts academic achievement in June (or even in the following September) for the simple reason that it already surreptitiously contains in itself most of the criterion it is claiming to predict. In fact, if short-term correlation is all that is required it would be sensible to abolish intelligence tests and predict from today's school achievement to tomorrow's.

But the idea of introducing intelligence tests has surely much broader objectives. It aims to give psychological understanding of causes of backwardness or sources of high performance, to permit prediction of more remote performances in time and place, and to permit us to apply a psychological science via functional analytical concepts. The writer had a student who never went to school till 17, who was soon getting straight A's as a junior in college, he did splendidly on a culture-fair test at 17 but would have been excluded from college on his traditional intelligence test or achievement performance. The discrepancy of $g_c$ and $g_f$ is, however, mostly small in the school years, so long as we have homogeneity of school, cultural subgroup, and social status. It is in the adult field, e.g., in selecting for an intelligence-demanding learning opportunity among persons of different ages and distances from school activities, in different regions, that failure to measure $g_f$ and $g_c$ separately could lead to serious misdirection. How fair is the Miller Analogies to engineering students competing for graduate school positions against English majors, compared with a culture-fair test? It is well known that the norms of extensively standardized traditional intelligence tests are invalidated within 2 or 3 years of their production by changing (usually increased!) school and cultural education levels. Is this intergenerational "creep" of norms on crystallized intelligence tests avoided by culture-fair tests as preliminary data (Cattell, 1957b) indicates? These are the problems which applied research should be facing.

In terms of basic research a realm of fascinating concepts and possibilities is opened up. Is it possible that $g_f$ minus $g_c$ might be the best predictor of organic brain damage? Can $g_f$ and $g_c$ also be separated by intraindividual, *P* technique factoring, demonstrating the functional unity of each in terms of susceptibility to day-to-day changing conditions and stimuli? What clarity might

---

fluid ability shows itself most highly, relative to crystallized ability, in complex relation and correlate eduction performances where crystallized ability is eliminated either by (*a*) all subjects being entirely untrained in the performance, or (*b*) all subjects being completely overlearnt in the fundaments (but not the relations, else intelligence would disappear)! Possibly Culture Fair Subtest Number 4, Topology, is an example of the former and Spatial ability of the latter, since the latter is a common requirement of bodily movement for people in all cultures. This definition of one property of the fluid general ability needs to be experimentally pursued in relation to the other hypothesized properties above, now that the factor can be isolated.

be brought to the vexed question of preschool ability organization by these concepts? (There are important leads here in the work of Burt, 1955; Hebb, 1942; Hofstaetter, 1954; McNemar, 1940; Reitan, 1958.) Will nature-nurture ratios calculated on $g_f$ measures prove to be altogether more highly genetically determined than $g_c$ measures? Is the $g_f$ IQ more constant?

Is it possible that what Thurstone called primary ability factors are psychologically the same as what we have called aids above? Probably both will consistently appear as first-order factors, relative to $g_c$ and $g_f$ as second order, and, although some may well be genetic in origin, the theory is worth exploring that many first orders will prove to be growths from a specific breakthrough (in the form of hitting upon a habit of analysis which becomes instrumental to a whole sequence of skill). In view of the elaborate writing of Piaget, and still more the speculations of his followers, regarding structures of this kind (see the recent summary by Hunt, 1961) it is high time that more precise correlational methods should attempt to demonstrate these patterns and relate them to the three types of formation here discussed.[14]

But perhaps good research strategy indicates that the first necessity today is the checking of the existence of the two general factors in groups of differing ages, homogeneities, and cultural settings, in relation to the obvious differences which would be expected from the theory. For example, this experiment needs repeating with age-homogeneous groups at 30 and 50 years of age; with adults heterogeneous as to age; and with mixed "normals" and known brain damaged cases. It needs to be tried with larger ranges of variables, identically used with two or three different cultural groups, to see whether the third-order achievement general factor can be separated both from the second-order crystallized intelligence and fluid intelligence factors, and especially we need age plots of the differing course of these two factors, each with and without speed limits in testing.

## SUMMARY

The theory of fluid and crystallized general intelligence factors has been examined in regard to an array of implications. A distinction has been drawn between a general achievement factor and the crystallized intelligence factor. Both, as distinct from fluid intelligence, will so change their pattern (*a*) with age, and (*b*) with culture as to constitute relatively poor predictors of other performances in groups of different ages and heterogeneous subcultures.

An experiment factoring 44 hypotheses-relevant variables measured on 277 seventh and eighth grade boys and girls, with sufficient noncognitive variables to permit effective rotation of any general cognitive factors which might appear, demonstrated the existence of two general ability factors. One fits the crystallized ability factor measured in traditional intelligence tests and the other a fluid general ability measured in culture-fair intelligence tests. Evidence is offered that the latter is neither a spatial ability factor nor the so-called "practical ability" (k) factor, but basic general intelligence. These two general abilities appear in a single third-order factor hypothesized to express the "formative fluid ability" partly responsible for the present level of both of them.

---

[14] In so far as the discovery of each aid (instrument for further problem solving) must be partly a function of fluid ability level, a part of the variance of all aid developments will be absorbed in the general factor of crystallized ability. The $F_A$'s of Equation 3 would represent, therefore, only the narrow group factor variance left over after this absorption in the general factor.

## BIBLIOGRAPHY

ANASTASI, A. The influence of specific experience upon mental organization. *Genet. Psychol. Monogr.*, 1936, **18**, 245–355.

BARGMANN, R. *A demonstration study of the effectiveness of factor analytical models.* Frankfurt: Hochschule fuer Internationale Paedagogische Forschung, 1955.

BLEWETT, D. Twin studies on the inheritance of intelligence and general neuroticism. Unpublished doctoral dissertation, University of London Library, 1954. (a)

BLEWETT, D. B. An experimental study of the inheritance of intelligence. *J. ment. Sci.*, 1954, **100**, 54–69. (b)

BRAINE, M. D. S. The ontogeny of certain logical operations: Piaget's formulation examined by nonverbal methods. *Psychol. Monogr.*, 1959, **73**(5, Whole No. 475).

BURKS, B. S. The relative influence of nature and nurture upon mental development. *Yearbook Nat. Soc. Stud. Educ.*, 1928, **27**, 219–236.

BURT, C. L. The evidence for the concept of intelligence. *Brit. J. educ. Psychol.*, 1955, **25**, 158–177.

CARRIGAN, P. M. Extraversion-introversion as a dimension of personality: A re-appraisal. *Psychol. Bull.*, 1960, **57**, 329–360.

CARROLL, J. B. The factorial representation of mental ability and academic achievement. *Educ. psychol. Measmt.*, 1943, **4**, 307–332.

CATTELL, R. B. The Cattell Group Intelligence Tests: Scales 0, 1, 2 and 3. London: Harrap, 1931.

CATTELL, R. B. *The fight for our national intelligence.* London: P. S. King, 1937.

CATTELL, R. B. A culture-free intelligence test: Part I. *J. educ. Psychol.*, 1940, **31**, 161–179.

CATTELL, R. B. Some theoretical issues in adult intelligence testing. *Psychol. Bull.*, 1941, **38**, 592.

CATTELL, R. B. The measurement of adult intelligence. *Psychol. Bull.*, 1943, **3**, 153–193.

CATTELL, R. B. The cultural functions of social stratification: I. Regarding the genetic basis of society. *J. soc. Psychol.*, 1945, **21**, 3–23.

CATTELL, R. B. *Personality: A systematic theoretical and factual study.* New York: McGraw-Hill, 1950.

CATTELL, R. B. *Factor analysis.* New York: Harper, 1952.

CATTELL, R. B. A note on Dr. Sloan's evidence regarding the value of culture free intelligence tests. *Amer. J. ment. Defic.*, 1955, **59**, 504–506.

CATTELL, R. B. *Personality and motivation structure and measurement.* New York: Harcourt, Brace & World, 1957. (a)

CATTELL, R. B. *The IPAT Culture Fair Intelligence Scales.* Champaign, Ill.: Institute for Personality and Ability Testing, 1957. (b)

CATTELL, R. B. A universal index for psychological factors. *Psychologia,* 1957, **1**, 74–85. (c)

CATTELL, R. B. The basis of recognition and interpretation of factors. *Educ. psychol., Measmt.*, 1962, in press.

CATTELL, R. B., & BELOFF, H. *The High School Personality Questionnaire.* Champaign, Ill.: Institute for Personality and Ability Testing, 1959.

CATTELL, R. B., & DICKMAN, K. A dynamic model of physical influences demonstrating the necessity of oblique simple structure. *Psychol. Bull.*, 1962, **59**, 389–400.

CATTELL, R. B., STICE, G. F., & KRISTY, N. F. A first approximation to nature-nurture ratios for eleven primary personality factors in objective tests. *J. abnorm. soc. Psychol.*, 1957, **54**, 143–159.

CATTELL, R. B., & SULLIVAN, W. The scientific nature of factors: A demonstration by cups of coffee. *Behav. Sci.*, 1962, **7**, 184–193.

CATTELL, R. B., & WARBURTON, F. W. *Objective personality and motivation tests: A theoretical introduction and practical compendium.* Urbana: Univer. Illinois Press, 1963.

COHEN, J. The factorial structure of the WAIS between early adulthood and old age. *J. consult. Psychol.*, 1957, **21**, 283–290.

CONNOR, D. The prediction of educational achievement from objective motivation factor measurement. Paper read at Midwestern Psychological Association Annual Meeting, Chicago, May 1961.

EL KOUSSY, A. H. A note on the Greys Analogy Test. *Brit. J. educ. Psychol.*, 1938, **4**, 294.

FERGUSON, G. A. On transfer and the abilities of man. *Canad. J. Psychol.*, 1956, **10**, 121–131.

FISHER, R. A. *The genetical theory of natural selection.* (2nd rev. ed.) New York: Dover, 1958.

FLEISHMAN, E. A., & HEMPEL, W. E. Changes in factor structure of a complex psychomotor task as a function of practice. *Psychometrika,* 1954, **19**, 239–252.

FORGUS, R. H. Early visual and motor experience as determiners of complex maze learning ability under complex and rich and reduced stimulation. *J. comp. physiol. Psychol.*, 1955, **48**, 215–220.

FORTES, M. A study of cognitive error. *Brit. J. educ. Psychol.*, 1932, **2**, 297–318.

FRENCH, J. W. The description of aptitude and achievement tests in terms of rotated factors. Psychological Monograph No. 5, 1951, University of Chicago Press.

GREENE, E. B. An analysis of random and systematic changes with practice. *Psychometrika*, 1953, **8**, 37–52.

GUILFORD, J. P., CHRISTENSEN, P. R., BOND, M. A., & SUTTON, M. A. A factor analysis study of human interests. *Psychol. Monogr.*, 1954, **68**(4, Whole No. 375).

HARLOW, H. F. Motivation as a factor in the acquisition of new responses. In, *Current theory and research in motivation: A symposium.* Lincoln: Univer. Nebraska Press, 1953. Pp. 24–49.

HAYES, K. J. Genes, drives and intellect. *Psychol. Rep.*, 1962, **10**, 299–342.

HEBB, D. O. The effect of early and late brain injury upon test scores, and the nature of normal adult intelligence. *Proc. Amer. Phil. Soc.*, 1942, **85**, 275–292.

HOFSTAETTER, P. R. The changing composition of intelligence: A study in T-technique. *J. genet. Psychol.*, 1954, **85**, 159–164.

HUMPHREYS, L. G. Discussion of Dr. Ferguson's paper. In P. H. Dubois, W. H. Manning, & C. F. Spies (Eds), Factor analysis and related techniques in the study of learning. Technical Report No. 7, 1959, Contract No. Nonr-816(02), Office of Naval Research. Pp. 183–187.

HUMPHREYS, L. G. The organization of human abilities. *Amer. Psychologist*, 1962, **17**, 475–483.

HUNT, J. McV. *Intelligence and experience.* New York: Ronald, 1961.

LANSDELL, H. Laterality of verbal intelligence in the brain. *Science*, 1962, **135**, 922–923.

LASHLEY, K. S. Cerebral organization and behavior. In K. S. Lashley (Ed.), *The neuropsychology of Lashley.* New York: McGraw-Hill, 1960.

LINE, W. The growth of visual perception in children. *Brit. J. Psychol. monogr. Suppl.*, 1931, No. 15.

LORD, F. M. The measurement of growth. *Educ. psychol. Measmt.*, 1956, **16**, 421–437.

LORD, F. M. Further problems in measurement of growth. *Educ. psychol. Measmt.*, 1958, **18**, 437–451.

MCNEMAR, Q. A critical examination of the University of Iowa studies of environmental influences upon the IQ. *Psychol. Bull.*, 1940, **37**, 63–92.

NEWLAND, T. E. The assessment of exceptional children. In W. M. Cruickshank (Ed.), *Psychology of exceptional children and youth.* New York: Prentice-Hall, 1962. Ch. 2.

PENROSE, L. F., & RAVEN, J. G. A new series of perceptual test: Preliminary communication. *Brit. J. med. Psychol.*, 1936, **16**, 28–30.

PIAGET, J. *The psychology of intelligence.* London: Routledge & Kegan Paul, 1947.

PRIBRAM, K. H. A review of theory in physiological psychology. *Annu. Rev. Psychol.*, 1960, **11**, 1–40.

REITEN, R. M. Qualitative versus quantitative mental changes following brain damage. *J. Psychol.*, 1958, **46**, 339–346.

SISK, H. L. A multiple factor analysis of mental abilities in the freshman engineering curriculum. *J. Psychol.*, 1940, **9**, 165–177.

SPEARMAN, C. *The abilities of man.* London: Macmillan, 1932.

SWENEY, A. B., & CATTELL, R. B. Dynamic factors in twelve year old children as revealed in measures of integrated motivation. *J. clin. Psychol.*, 1961, **17**, 360–369.

THURSTONE, L. L. *Primary mental abilities.* Chicago: Univer. Chicago Press, 1938.

THURSTONE, L. L. *Multiple factor analysis.* Chicago: Univer. Chicago Press, 1947.

TUCKER, L. R. Determination of generalized learning curves by factor analysis. In P. H. Dubois, W. H. Manning, & C. F. Spies (Eds.), Factor analysis and related techniques in the study of learning. Technical Report No. 7, 1959, Contract No. Nonr-816(02), Office of Naval Research.

VERNON, P. E. *The structure of human abilities.* New York: Wiley, 1950.

WOODROW, H. The relation between abilities and improvement with practice. *J. educ. Psychol.*, 1938, **29**, 215–230.

WOODROW, H. Factors in improvement with practice. *J. Psychol.*, 1939, **7**, 55–70.

(Received April 20, 1962)

*Journal of Educational Psychology*
1963, Vol. 54, No. 1, 23–32

# RELATIONSHIPS BETWEEN TEACHER NEEDS AND TEACHER BEHAVIOR IN THE CLASSROOM[1]

NORMAN E. WALLEN, ROBERT M. W. TRAVERS, IAN E. REID,
AND KENNETH H. WODTKE
*University of Utah*

2 samples of elementary school teachers (N = 77 and 41) were studied with respect to consistencies between classroom behavior and "needs" as measured by a paper-and-pencil questionnaire. The paper-and-pencil device provided scores on Achievement, Affiliation, Recognition, and Control needs. Behavioral data consisted of observations recorded as ratings and also as Q sorts and an analysis of verbal statements made by the teachers. The principal findings were: (a) significant positive relationships among the various measures of teacher behavior on the Control and Affiliation dimensions, (b) a significant negative relationship between Control and Affiliation behavior as viewed by the observers, and (c) significant positive correlations (.30–.50) between Control need as measured by the questionnaire and controlling behavior of the teachers.

In recent years, many writers have expressed the view that major improvements in education must await further knowledge as to the actual classroom behavior of teachers and its consequences for student behavior. Notable contributions to this problem have been made by Ryans (1960), Medley and Mitzel (1959), Cogan (1958), Washburne and Heil (1960), and others. An important aspect of the general problem is the extent to which teacher behavior is measurable and predictable. The study here reported is an attempt to delineate dimensions of classroom behavior of elementary school teachers and to assess their predictability by paper-and-pencil measures of need.

## SAMPLE

Two samples of teachers were involved in the study. The first sample consisted of 77 teachers in Grades K through 6 representing five schools in Salt Lake City. In selecting these schools attention was given to securing a variety of socioeconomic backgrounds. The second sample consisted of 41 teachers representing two schools in a suburban, semirural community. There are differences between the two samples as indicated by different patterns of intercorrelations among some predictor and behavioral measures used in the study and by the fact that there is a greater turnover in personnel in the Salt Lake City schools than in the suburban schools. The teachers varied across the entire range of age and experience. Good cooperation was obtained. Only four teachers refused to admit the project staff to their classrooms, though with one exception they were friendly in other respects. On occasion teachers would indicate that they would rather the observer return at another time but these incidents were infrequent.

## MEASURES OF TEACHER BEHAVIOR

Our primary interest was in the following postulated dimensions of teacher behavior:

1. *Achievement*: degree of emphasis on academic activities and the importance of excellence
2. *Affiliation*: degree of interaction with students and emphasis on warm personal relationships
3. *Control*: degree to which the moment to moment behavior of students is controlled by the teacher
4. *Recognition*: degree to which the teacher is the center of attention

Two means of assessing these dimensions were used. The first was based on sampling of the verbal behavior of the teacher; the second consisted of observer ratings and elab-

---

[1] The research reported herein was supported through the Cooperative Research Program of the Office of Education, United States Department of Health, Education, and Welfare.

orations thereof. Each of the participating teachers was visited in class by two members of the research team. After preliminary visits to reduce the effect of an observer, data collection was begun. Each teacher was visited at least three times by each observer. Three devices were used as follows.

*Adaptation of the technique developed by Withall (1949).* This technique provides a sample of the teacher's verbal behavior by recording verbatim the first utterance made during consecutive 1-minute periods. Each observer collected two samples of 25 statements each, resulting in a total of 100 statements for each teacher over a minimum of four observation periods.

The statements were subsequently classified into a system containing 26 categories. Our procedure was to have three persons score each teacher statement; the category assigned by at least two of the three scorers was accepted. This criterion was achieved for 94% of the statements. Categories were then combined to provide scores for Achievement, Affiliation, and Control behaviors and for a dimension reflecting activities related to classroom Management. The Recognition dimension was not appropriate for these data. The reliability of these measures was assessed by means of Lindquist's (1953) analysis of variance technique with results as follows: Achievement, .50; Affiliation, .54; Control, .56; and Management, .47. These reliabilities are somewhat higher than those obtained by Medley and Mitzel (1949) for the six Withall categories.

*Ratings made by the observers.* In addition to ratings made on the major dimensions cited earlier (Achievement, Affiliation, Recognition, and Control), ratings were made on 31 other dimensions gleaned from prior research, including amount of time spent alone in class, vigor, confidence, etc. These latter ratings are not discussed further here except as they relate to the third measurement approach as discussed below. A seven-point scale was used. A point of interest is that although the correlation between observers for a particular characteristic ranged from .15 to .58, even in cases of low agreement summary scores based on two raters were significantly and predictably related to other measures in more instances than would occur by chance. The reason may be that high and low summary scores reflect a characteristic strong enough in a particular teacher to be agreed upon by the raters; cases of disagreement receive summary scores in the middle of the distribution. Thus a summary score may have considerable validity.

*Factor scores derived from* Q *sorts.* In addition to independent ratings, a *Q* sort procedure was used. This procedure has several advantages over ratings. First, it permits the systematic comparison of each teacher with every other with respect to each characteristic, hence providing a more stable frame of reference. Second, it permits the observers to arrive at a consensual evaluation based on discussion as opposed to an additive pooling of their individual assessments.

The disadvantages of this procedure are that the assessments are further removed in time from the behavior of the teachers and that the consensual assessment may introduce error resulting from the personal interaction of the observers.

The following procedure was followed for each of the two samples of teachers. A deck of cards was prepared with one card for each teacher. Photographs were taken of all of the teachers who would consent and the pictures were pasted on the cards. The purpose of this was to provide a reminder of each teacher for the raters. Pictures were obtained for 60 of the 77 teachers in Sample 1 and 39 of the 41 teachers in Sample 2. Each of the 24 traits was then discussed with respect to particular teachers before an attempt was made to sort the entire batch of cards into the prescribed categories.

The sort was conducted to provide the following proportions of cases over seven categories: .07, .12, .18, .24, .18, .12, .07. The procedure was as follows: As each card was turned, each sorter independently assigned it a numerical value corresponding to one of the seven piles. If there was disagreement between the two sorters, the final placement was decided after discussion. The sorters were impressed with a rather high degree of agreement though observer correlations were not computed. Following this process, cards were shuffled up or down in the sequence of piles, by mutual agreement, until the correct number was obtained in each pile.

In performing these *Q* sorts, the research workers were concerned only with behavior of the teachers within the classroom. This is a point to be emphasized since in some cases there seemed to be striking differences between the behavior of the teacher in the classroom and behavior in other situations in which these teachers could be observed. In addition, an attempt was made to consider the behavior of the teacher in relation to that of the child. If, for example, the degree of hostility of the teacher was to be judged then the judgment was made as to whether the pupils reacted as though the teacher were hostile. This problem

TABLE 1

ROTATED FACTOR MATRIX—VARIMAX SOLUTION—Q SORT DATA

| Variable | Factor | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| 1. Aloof—Response | **.81** | .02 | −.25 | −.10 | .14 |
| 2. Harsh—Kindly | **.85** | .18 | −.25 | −.22 | −.02 |
| 3. Dull—Stimulating | .46 | **−.63** | −.33 | .13 | −.24 |
| 4. Excitable—Controlled | .26 | **.77** | −.33 | −.06 | .16 |
| 5. Uncertain—Confident | .35 | −.14 | **−.73** | .08 | −.21 |
| 6. Disorganized—Systematic | −.13 | −.08 | −.49 | .08 | **−.52** |
| 7. Inflexible—Adaptable | **.88** | .01 | −.21 | .10 | .11 |
| 8. Friendly—Hostile | **−.84** | −.02 | .22 | .18 | .07 |
| 9. Much direct control—Little direct control | .64 | .27 | .06 | .06 | .40 |
| 10. Emphasized learning—Lack of concern for learning | .08 | .28 | .10 | −.22 | .87 |
| 11. Jokes, laughs—Serious | −.08 | .27 | .21 | −.05 | .06 |
| 12. Frequently punishes—Gives support | .63 | .49 | −.31 | −.11 | .10 |
| 13. Much emotional evidence of frustration—Little evidence of emotional frustration | .49 | .27 | **−.73** | −.06 | −.03 |
| 14. Vigorous—Quiet | −.05 | **.86** | .17 | .00 | .21 |
| 15. Functions often as a source of knowledge—Functions seldom as a source of knowledge | .01 | .49 | .14 | .12 | .54 |
| 16. Delegates much authority—Delegates little authority | −.74 | −.15 | .05 | −.23 | −.01 |
| 17. Spends much time working alone—Spends little time working alone | .18 | −.13 | −.05 | **−.69** | .18 |
| 18. Negativistic—Positive | .72 | −.16 | −.51 | .00 | −.20 |
| 19. Permits great variety of activities, noise etc.—Activities very orderly | **−.86** | −.06 | .01 | .03 | −.26 |
| 20. Very verbal—Little verbal behavior | .06 | **.80** | −.06 | .35 | .22 |
| 21. Low achievement motivation—High achievement motivation | −.03 | −.43 | −.02 | .08 | **−.75** |
| 22. Low affiliation motivation—High affiliation motivation | **.87** | .04 | −.18 | −.26 | −.01 |
| 23. Low recognition motivation—High recognition motivation | −.11 | **−.71** | −.02 | −.12 | −.29 |
| 24. Neuroticism—High ego strength | .43 | −.02 | **−.76** | −.18 | .01 |

Note.—Boldface values indicate variables used in assigning factor scores to teachers.

was well illustrated in the case of one teacher whose remarks might well have been interpreted by a casual observer as hostile but which the children obviously did not interpret in this way. This latter factor introduced an additional element into the Q sort situation which is likely to lower the reliability of the procedure, but it seemed to be justified.

In order to achieve a more parsimonious treatment of these variables a centroid factor analysis with varimax rotation was performed for the first sample of teachers ($N = 77$). The rotated matrix is shown in Table 1. This analysis resulted in five factors labeled as follows:

1. Cold, controlling versus warm, permissive
2. Vigorous, dynamic versus dull, quiet
3. Insecure, anxious versus confident
4. Spends much time alone versus little time alone
5. Much academic activity versus little academic emphasis

These dimensions are quite similar to those of Ryans (1960). It is interesting to note that Affiliation and Control do not emerge as separate dimensions, a result supported by the correlation of the ratings on these variables. As viewed by the observers there is a distinct tendency for the controlling teacher to be cold and aloof though there are exceptions.

Each teacher was assigned a score on each factor which was the sum of the *Q* sort ratings on the dimensions having highest loading on the factors. These loadings are in boldface type in Table 1.

Intercorrelations among the *Q* sort factors are shown in Table 2. The first point to be noted is that there are substantial correlations between certain of the factor scores. Thus, in spite of the fact that the factor analysis maintained orthogonal factors, these independent factors tend in some cases, to produce correlated *factor scores* as applied to this sample of teachers. The nature of the correlations among factor scores varies considerably from Sample 1 to Sample 2. Thus, for Sample 1 the correlation between scores on Factors 1 and 3 indicates that the teacher who was viewed as warm and permissive also tended to be viewed as nonanxious and emotionally stable. The correlation between scores on Factors 2 and 5 indicates that teachers viewed as stimulating appear more concerned with school learning as such. These correlations, however, are much lower in Sample 2. In Sample 2, the magnitude of correlations shifts such that Factor 1 (warm, permissive versus cold, controlling) scores correlate more highly with Factor 4 (little time alone) scores than Factor 3 (ego strength) scores.

Since there are no other criteria (other than ratings) which are logically related to Factors 2 and 3, the possibility exists that these factors themselves as well as subsequent factor scores reflect a bias common to the observers rather than a description of behavior. Factors 1 and 5 are on more solid ground since they are related to an objective criterion—teacher statements—as will be discussed next.

Intercorrelations among the various measures of classroom behavior are shown in Table 3. Those instances in which the correlation was nonsignificant (.05 level) for *both* samples have been omitted.[2] The major finding is that the measures of Control and Affiliation are

TABLE 2

INTERCORRELATIONS AMONG FACTOR SCORES FOR THE TWO SAMPLES OF TEACHERS

| Factor | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 1. Warm, permissive—Cold, controlling | .07<br>(−.16) | .62**<br>(.34)* | .27*<br>(.62)** | .12<br>(.28) |
| 2. Quiet, controlled, dull—Stimulating, active | | .00<br>(−.04) | −.12<br>(−.31)* | .55**<br>(.28) |
| 3. High ego strength—Low ego strength | | | .13<br>(.02) | −.23*<br>(−.31)* |
| 4. Spends much time alone—Little time alone | | | | .12<br>(.20) |
| 5. Little concern with academic achievement—Much concern | | | | |

Note.—Correlations for Sample 2 are in parentheses.
* $p \leq .05$.
** $p \leq .01$.

---

[2] Two-tailed tests are reported throughout for ease of communication although one-tailed tests are justified in many instances wherein a positive correlation between different measures of the same characteristic or between a measure of need and its hypothesized counterpart in behavior were, of course, predicted.

TABLE 3

CORRELATIONS AMONG MEASURES OF TEACHER BEHAVIOR

| Factor | Teacher statements | | | | Rating | | | |
|---|---|---|---|---|---|---|---|---|
| | Achievement | Affiliation | Control | Management | Achievement | Affiliation | Recognition | Control |
| 1. Warm—Cold | | .26*<br>(.38)* | −.39*<br>(−.50)** | .26*<br>(.01) | | .62**<br>( .68)** | | −.54**<br>(−.62)** |
| 2. Quiet—Stimulating | | | −.22*<br>(−.16) | | −.46**<br>(−.24) | | −.77**<br>(−.55)** | −.45**<br>(−.20) |
| 3. High—Low ego strength | | | −.42**<br>(−.15) | | .14<br>( .30)** | .38**<br>( .02) | | −.34**<br>(−.01) |
| 4. Much—Little time alone | | | | | | .31**<br>( .63)** | | .02<br>(−.47)** |
| 5. Little—Much academic emphasis | −.25*<br>(−.09) | | −.02<br>(−.33)* | | −.73**<br>(−.60)** | | −.44**<br>(−.33)** | −.34**<br>(−.51)** |
| Teacher statements Achievement | | | | | | | | |
| Teacher statements Affiliation | | | | | −.45**<br>(−.09) | .33**<br>( .41)** | .22*<br>(−.02) | −.33**<br>(−.30)* |
| Teacher statements Control | | | | | | −.30**<br>(−.31)* | .22*<br>( .05) | .43**<br>( .28) |
| Teacher statements Management | | | | | | | | −.06<br>(−.34)* |

Note.—Sample 2 values are in parentheses.
* $p \leq .05$.
** $p \leq .01$.

related in a consistent fashion across the three instruments. The correlations between the teacher statement measure and the ratings are .33 and (.41)[3] for Affiliation, and .43 and (.28) for Control. Factor 1 which is a combination of the Affiliation and Control dimensions correlates .26 and (.38) with the teacher statement Affiliation score; .62 and (.68) with rating measures of Affiliation; −.39 and (−.50) with the teacher statement Control score; and −.54 and (−.62) with rating measures of Control. These results are consistent with the negative correlation between Affiliation and Control as reported previously.

The remaining correlations between factor scores and *ratings* are, in general, as expected though it must be kept in mind that these do not constitute independent assessment techniques since both are based on observer classifications of teacher characteristics. The correlations which are in accord with expectations are: −.77 and (−.55) between Factor 2 (dull) and rating on Recognition; .31 and (.63) between Factor 4 (little time spent alone) and rating on Affiliation; −.73 and (−.60) between Factor 5 (nonachievement oriented) and rating on Achievement. The correlations between ratings on Control and Affiliation with Factor 3 (high ego strength) are −.34 and .38 for Sample 1 and are consistent with other relationships found between these variables. The correlations between Factor 2 (dull) and ratings on Achievement and Control (−.46 and −.45, respectively, for Sample 1) are to be expected since this factor is apparently primarily a "recognition" factor and the correlations among ratings on Achievement, Recognition, and Control are all substantial for Sample 1. The same explanation applies to the correlation of −.44 (−.33) and −.34 (−.51) between Recognition and Control ratings and Factor 5 which is essentially an "achievement orientation" factor. These correlations differ in Sample 2 since the correlations among factors and among ratings changed somewhat from Sample 1 to Sample 2. Thus, in summary, the ratings and factors show the internal consistency one should expect since they are two approaches based on the same original information: observer impressions.

---

[3] The first value pertains to Sample 1. The second value, in parentheses, pertains to Sample 2.

As mentioned previously the teacher statements measures provide support for the observational measures of Affiliation and Control dimensions and combinations thereof. With respect to the Recognition and Achievement dimensions, there is less correspondence. The correlations between teacher statements Affiliation and Achievement rating in Sample 1, −.45, and between teacher statements Management and Recognition rating in Sample 2, (.34), are close to zero in the other sample. The teacher statements Achievement measure provides only slight support for the observational measures of achievement orientation since the correlations are .11 and (.21) with Achievement rating and −.25 and (−.09) with Factor 5.

## NEED MEASURE

The measure of need used was the Teacher Preference Schedule (TPS) developed by Stern and Masling (1958).[4] It consists of 100 statements relating directly to teaching or to children to which the respondent indicates the extent of her agreement. In addition to the 10 needs described by the authors, the test was scored for Achievement, Affiliation, Control, and Recognition need according to keys prepared by the present writers. Split-half reliabilities of these measures are not reported by the authors; however, they were computed in the present study for the two most promising needs (as discussed below) with the following results: Control need = .80 and Affiliation need = .64. Inasmuch as we were primarily interested in the needs of Achievement, Affiliation, Control, and Recognition and since these scores proved most predictive of teacher behavior, subsequent discussion is restricted to these scores.

Intercorrelations among these need scores are shown in Table 4 and indicate that while the scores are quite unrelated for Sample 1, the Achievement score correlates quite substantially with the Affiliation and Recognition scores in Sample 2. The only correlation which is consistent for both samples is that between Affiliation and Recognition.

## RELATIONSHIPS BETWEEN NEED SCORES AND BEHAVIOR

The results are shown in Table 5. Those instances in which the correlation was nonsignificant (.05 level) for both samples have, again, been omitted.

TABLE 4

CORRELATIONS AMONG SCORES ON THE TEACHER PREFERENCE SCHEDULE

| TPS | TPS<br>Affiliation | <br>Recognition | <br>Control |
|---|---|---|---|
| Achievement | .05<br>(.48)** | .19<br>(.52)** | −.31**<br>( .20) |
| Affiliation | | .39**<br>(.60)** | −.07<br>(−.22) |
| Recognition | | | .21<br>(−.15) |

Note.—Sample 2 values are in parentheses.
* $p \leq .05$.
** $p \leq .01$.

---

[4] Several other paper-and-pencil devices were used to assess needs including a projective device patterned after French's (1955) A Test of Insight and devices requiring selection of preferred ways of responding to classroom situations and statement of preferences regarding educational objectives. None of these devices predicted classroom behavior though the projective test has demonstrated considerable validity as a predictor of performance in laboratory studies.

TABLE 5

RELATIONSHIPS BETWEEN TPS AND MEASURES OF CLASSROOM BEHAVIOR
(Factor scores, teacher statements, and ratings)

| TPS | Factor score | | | | | Teacher statements | | | | Ratings | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | Ach. | Aff. | Con. | Man. | Ach. | Aff. | Rec. | Con. |
| Achievement | | | | −.20<br>(.34)* | −.12<br>(.34)* | | −.08<br>(.40)** | | | .19<br>(−.36)* | −.18<br>(.31)* | | |
| Affiliation | | **.24***<br>(−.01) | | −.02<br>(.42)** | .18<br>(.35)* | | | | | −.12<br>(−.36)* | **.25***<br>(.30)* | **.34****<br>(−.20) | −.17<br>(−.34)* |
| Recognition | | | | .15<br>(.32)* | | **.22***<br>(−.04) | | | **−.38****<br>(.05) | | **.22***<br>(.39)* | | |
| Control | **−.48****<br>(−.47)** | | **−.33****<br>(−.24) | −.08<br>(−.44)** | | | | **.34****<br>(.29)* | | | **−.38****<br>(−.39)* | | **.51****<br>(.30)* |

Note.—Those values which were predicted and are in the predicted direction are in boldface type.
* $p \leq .05$.
** $p \leq .01$.

The primary result is the finding that the measures of controlling behavior are consistently and quite highly related to the appropriate scales of TPS. The most impressive relationships are between the Control need on TPS and the behavior rating for Control, .51 and (.30), and the teacher statements Control, .34 and (.29). In view of the negative correlation between Control rating and Affiliation rating, a negative correlation between Control need on TPS and Affiliation rating would be expected and does in fact occur, −.38 and (−.39). The relationship between Control on TPS and the factor scores is consistent with the preceding results in that both Factors 1 (warm, permissive) and 3 (confident), the scores of which are correlated one with the other, correlate significantly and negatively as expected with Control need on TPS. To summarize, the Control score on TPS appears to be a promising predictor of the principal dimension of teacher behavior which emerges from our study—i.e., warm, permissive versus cold, controlling. One should perhaps, emphasize again at this point that this dimension results from classroom *observation* measures. The best conclusion is that TPS Control score predicts controlling behavior on the part of the teacher and that, as viewed by the researchers, controlling behavior is highly related to lack of warmth and to a lesser degree to lack of confidence on the part of the teacher.

In general the remaining relationships support the above discussion in that TPS measures are related as expected to behavior measures. Those significant correlations which were in the predicted directions are underlined. Thus, the Affiliation score on TPS correlates positively, as expected, with the Affiliation rating. It does not, however, correlate as expected with the teacher statements Affiliation score.

With respect to the remaining measures of teacher behavior: Achievement orientation (rating, factor, and verbal statements), Recognition (rating and factor) and teacher statements Management, few TPS measures relate significantly to them nor does a meaningful pattern emerge. There are no consistent predictors either for teacher statements Achievement scores or for Recognition (factor and rating) or for teacher statements Management scores. It is important to note that correlations which were expected to be low and nonsignificant are, in general, in accord with expectation.

It should be noted that the Achievement dimension appears much more predictable for Sample 2 than for Sample 1. Whereas there is no systematic pattern of relationships in Sample 1, in Sample 2 both TPS Achievement and Affiliation scores are related to the Achievement Factor 5 and to Achievement ratings, such that high achievement behavior is related to low Affiliation and *low* Achievement need as measured by TPS. The negative correlation with TPS Achievement is contrary to expectation but consistent with the positive correlation between TPS Achievement and TPS Affiliation for this sample.

It is also of interest that Factor 4 (little time spent alone) is correlated, as would be expected, with TPS Affiliation in Sample 2. It is also correlated with TPS Achievement and Recognition scores and, negatively, wtih the Control score. The former two correlations are expected since there exist substantial correlations among TPS scores for this sample. The relationship with Control cannot be explained on this basis but is consistent with expectation. Since these results obtain only for Sample 2, they are merely suggestive.

## DISCUSSION

These results seem quite encouraging with respect to the predictability of at least some aspects of teacher behavior in the classroom. It was disappointing to find so little predictability for the Achievement or academic emphasis dimension. Perhaps this indicates that this aspect of classroom teacher behavior is less a function of teacher personality than the Control-Affiliation aspect; perhaps being more affected by external pressures such as attitudes of the principal or immediate situational conditions.

It is of considerable interest to note that the only effective predictor of teacher behavior in the study was a relatively straightforward questionnaire. As noted earlier, several other instruments were used with no success including a projective test which had shown considerable validity in several "laboratory" studies conducted by us and whose predecessors have a generally, though not unequivocally, favorable record (Atkinson, 1958; Travers, Wallen, Reid, & Wodtke, 1961). It may be of value to consider possible explanations for this.

First, there is the fact that situations in which instruments similar to our projective test of needs have been used most successfully have been ones in which the cues related to motive arousal have been under the control of the experimenter. In addition, the best success has been achieved with instruments which have been administered also under conditions which permit some control over the cues present while the instrument was being taken. From a theoretical standpoint prediction should be best when control is exercised over the motive aroused both in the testing situation and in the criterion situation. Such control generally cannot be exercised when projective devices are used for making predictions in naturally occurring situations. Certainly, in the present research, the practical impossibility of maintaining such control of both the testing and teaching conditions would reduce the possibility of making predictions.

This point may be made in a slightly different way. If one desires to predict typical performance of a person under stimulus conditions which are familiar to him, the best prediction may very well result from devices which ask him, in effect, how he behaves in such situations. Such are the findings of the present study and they are similar to results appearing elsewhere in the literature. It may, in fact, be a misnomer to label such measures "needs." Rather, such scores may perhaps better be viewed simply as self-descriptions of preferences, behavior, and the like without becoming involved in the supplementary meanings which attach to the term need.

The contribution of more subtle devices such as projective techniques, which attempt to assess dynamic as opposed to behavioral variables, should be toward the prediction of behavior under stimulus conditions which are novel to the person. The development of so-called dynamic psychology revolves on the premise that individuals having differing need structure *may* exhibit similar behavior and that individuals with similar need structure *may* show very different behavior. The value of need assessment, then is primarily in predicting behavior under novel conditions wherein well established behaviors are less likely to occur. It may be noted that laboratory studies wherein projective measures have the most predictive success meet the criterion of novelty to a much greater extent than classroom situations as studied in this project.

Second, there is the possibility that persons who take projective tests in situations other than the laboratory or where they suspect that the test is given for reasons other than academic ones, may manifest response sets which seriously detract from the significance of the scores. In the present research, evidence indicates that the responses of teachers to the projective test were different from those of students to whom it was administered as part of a general class requirement. While the scores of the student population showed that the four scales had a high degree of independence, the scores of the teachers showed a high degree of relationship among the scales. Several explanations may be suggested for this phenomenon. One is that the teachers may differ from the students by representing a population in which there are relatively highly correlated needs. Another is that the teachers may differ one from another in a variable which might be described as imagery fluency. Thus some teachers may obtain low scores on this device because they fail to respond as profusely to any of the cues provided by the items while others may respond profusely to all of the items.

The problems raised here are persistent in the use of projective devices and are not unique to this study. While most measuring instruments can be used in situations other than those in which they have been demonstrated to have validity, this generalization may not apply to projective devices because of their extreme sensitivity to testing conditions.

## REFERENCES

ATKINSON, J. W. (Ed.) *Motives in fantasy, action, and society*. New York: Van Nostrand, 1958.

COGAN, M. L. The behavior of teachers and the productive behavior of their pupils. *J. exp. Educ.*, 1958, **27**, 89–124.

FRENCH, E. G. Some characteristics of achievement motivation. *J. exp. Psychol.*, 1955, **50**, 232–236.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. New York: Houghton Mifflin, 1953.

MEDLEY, D. M., & MITZEL, H. E. Application of analysis of variance to the estimation of the reliability of observations of teachers' classroom behavior. *J. exp. Educ.*, 1949, **17**, 347–361.

MEDLEY, D. M., & MITZEL, H. E. Some behavioral correlates of teacher effectiveness. *J. educ. Psychol.*, 1959, **50**, 239–246.

RYANS, D. G. *Characteristics of teachers.* Washington, D. C.: American Council on Education, 1960.

STERN, G. G., & MASLING, J. M. Unconscious factors in career motivation for teaching. Final Report, 1958, United States Office of Education, Contract No. SAE 6459.

TRAVERS, R. M. W., WALLEN, N. E., REID, I. E., & WODTKE, K. H. Measured needs of teachers and their behavior in the classroom. Final Report, 1961, United States Office of Education, Contract No. 444(8029).

WASHBURNE, C., & HEIL, L. M. What characteristics of teachers affect children's growth? *Sch. Rev.*, 1960, **68**, 420–427.

WITHALL, J. The development of a technique for the measurement of social-emotional climate in the classroom. *J. exp. Educ.*, 1949, **17**, 347–361.

(Received February 2, 1962)

*Journal of Educational Psychology*
1963, Vol. 54, No. 1, 33–37

# FAMILIAL CORRELATES OF ORIENTATION TOWARD FUTURE EMPLOYMENT AMONG COLLEGE WOMEN[1]

ALBERTA ENGVALL SIEGEL[2] AND ELIZABETH ANN CURTIS[3]
*Pennsylvania State University*

Information obtained from individual interviews with 43 young women, a random sample from the population of sophomore women on the campus of a large university, was coded by 2 independent scalers to yield measures of the work orientation of the Ss and 5 characteristics of their families: socioeconomic status, parents' educational level, mother's work orientation, parents' views on purpose of college, and parents' attitudes toward importance of education for the daughter. Only 1 of these latter characteristics—mother's work orientation—was found to be significantly correlated with work orientation in Ss. Most Ss stated they intended to marry and to work. The homogeneity of the sample may have operated to depress observed correlations.

This study bears on the often-noted discrepancy between young women's expectations and plans for their future, on the one hand, and the actual likelihoods facing them, on the other. Whereas adult women are increasingly entering the nation's working force, young women are said to be continuing to focus their expectations for the future upon home and family. The purpose of this study was to examine the extent to which college women are oriented toward future employment, and to determine the relation between students' orientation toward work and certain characteristics of their families.

Formulated under the assumption that the validity of individual interview responses, carefully coded by well-trained judges, is much higher than the validity of more easily obtained responses to a written questionnaire, this study stands in contrast to previous ones with similar purposes in its use of individual open-end interviews.

## METHOD

### *Subjects*

Subjects were native-born white women enrolled in the sophomore year on the University Park campus of Pennsylvania State University, a large land-grant institution, 90% of whose undergraduates are residents of Pennsylvania, the great majority of them from the upper one-fifth of the high school graduating class.

A random sample of 46 subjects was drawn from a list of all third and fourth semester women on the campus during the spring semester of 1961. Those selected were contacted by letter, and then by telephone call and a confirming letter. Subject response was excellent: 44 of the subjects completed the interview. One of these was dropped from the study because of foreign birth, so results are reported for $N = 43$.

From the interview data, it was determined that the majority of the subjects (84%) were 19 years old, the remaining few either 18 or 20. They represented a diversity of major curricula, with over one-half enrolled in various education majors (a proportion which is similar to that in the population of sophomore women).

---

[1] We wish to thank Lorraine B. Dennis and Miriam B. Haas for their assistance in the execution of this study, and W. L. Morgan and M. L. Ziegler for administrative cooperation. This investigation was supported in part by a Public Health Service Research Grant (M-3911) to the senior author from the National Institute of Mental Health, Public Health Service.

[2] On leave, now at the Center for Advanced Study in the Behavioral Sciences, Stanford, California.

[3] Now at Ohio State University.

The majority of subjects (58%) had fathers working in the top two levels of the seven-level scale of Hollingshead and Redlich (1958, pp. 390–391), and all but 9% had fathers in the top three levels. About one-half of the fathers were college graduates; only 15% had achieved less than high school graduation. Of the mothers, only 17% had achieved a college education, but an additional 39% had received some education beyond high school without attaining a baccalaureate degree. Only 11% of the mothers had achieved less than high school graduation. All but 14% of the subjects came from homes in which both parents were living. The homes of 54% were in suburbs; while 30% were in towns, 9% were in rural areas, and 7% were in cities.

## *Interview*

*Procedure.* All interviews were conducted in a small conference room by the same person, a young women. Interviews consisted, first, of a series of open-end questions concerning the subject's purpose in seeking a college education and her plans for the years after college, and, second, of a longer series of open-end questions concerning family background and parents' attitudes toward education. Interviews typically required 30 minutes.

*Coding.* Responses were coded directly from tape recordings, on the basis of codes devised in advance through pilot work.

Since the purpose was to relate work orientation of the subjects to certain family background variables, it was necessary to obtain scores on work orientation and the other variables which were independent of each other (not intercontaminated). Accordingly, two coders were used. Coder A listened to the initial series of questions and coded the responses for work orientation, using scales to be detailed below. Coder B listened only to the later series of questions, coding for the other variables in ignorance of the subject's responses relevant to work orientation.

To check the reliability of the codes, a third coder, C, was employed, who worked with a sample of 20 of the tapes (every second case). Coder C initially coded responses to the second series of questions, thereby checking the reliability of B, and then coded the responses to the first series, thereby checking the reliability of A.

In general, good intercoder agreement was obtained. For the 11 nominal (categorical) codes, the median reliability was 90% agreement, with a range from 70% to 100% agreement. For the 9 ordinal (rating scale) codes, the median reliability was rho = .93, with a range from rho = .64 to rho = .99. (For only 2 of the 9 ordinal codes did reliability fall below rho = .81.)

## *Scales*

*Work orientation of subjects.* The initial questions in the interview related to the subject's plans, expectations, and hopes concerning employment. These were arranged in an order so that the initial ones did not arouse the salience of the work versus home issue, while the later ones probed increasingly in this sometimes sensitive area. Responses to each series, described in order below, were coded as *work oriented* or *not work oriented*.

1. The subject was asked to identify her major field of study and to state the purpose of that curriculum. A response was coded work oriented if the subject described her curriculum in terms of preparation for an occupation.

2. The subject was then asked to state the role she thought her college education would play in her future, and to sketch her plans after leaving college. A response was coded work oriented if the subject mentioned employment in her plans for the near future.

3. The topic of marriage was then introduced explicitly: "When you think of the future, how does marriage figure in your plans?" A response was coded work oriented if the subject stated that she intended to continue working after marriage.

4. The possibility that a women might wish to stop working after marriage was then suggested, and the subject was asked how she felt about a woman's working after she had children. A response was coded work oriented if the subject stated she wanted to work throughout her marriage, or if she wanted to work until children were born and then to return to work when they entered school.

5. Finally, the subject was asked "Suppose you were thinking about marrying a particular man, and you realized that if you married this man it would be impossible for you to continue to work. Would you marry this man?" This question was deliberately phrased so that the nature of the husband-work conflict was ambiguous. Not a single subject stated that she would give up the man in order to retain her freedom to work, so a response was coded work oriented if the subject's reply was equivocal, e.g., indicated that such a choice could be avoided.

6. To supplement the foregoing codes, the

interviewer (who was also later Coder A) made an overall assessment of each subject's motivation for work achievement, using the entire content of the interview and her impressions as interviewer and coder. She used a five-point scale ranging from "strongly oriented to work, ambitious, interested in achievement through work" to "not at all oriented to work; work is not important to self satisfaction." The subjects who fell at the first three levels of the scale were called work oriented.

From the six codes which have been described, an overall index of work orientation was obtained, consisting of the sum of the subject's scores. With a possible range from 6 to 17 (low scores reflecting strong work orientation), obtained scores ranged from 7 to 16, with a median of 10.

*Family background.* The index for each subject's *socioeconomic background* was a weighted sum of her scores on father's education and father's occupation, based on the weights derived by Hollingshead and Redlich (1958, pp. 390–397).

Similarly, an index of *parent's educational level* was the sum of scores on the seven-level codes for mother's educational level and father's educational level (Hollingshead & Redlich, 1948, p. 391).

For an index of *mother's work orientation*, replies were used to a series of questions about the mother's work history and attitude toward working. First, the subject was asked "Is your mother working now?" Just over 40% of the replies were affirmative. Second, the subject was asked about her mother's attitude toward work—whether she would prefer to work or not to, irrespective of her present employment. The subjects indicated that 60% of the mothers preferred working. Third, the subject was queried in detail about her mother's work history, and from the replies each mother was categorized according to whether or not she had a consistent history of being a working mother. Only 21% of the mothers were "working mothers" in the sense of having worked fairly consistently after their children were born or after they entered school. Fourth, the nature of the mother's work was coded. For the purposes of the index of mother's work orientation, the professional workers (36%) and the white-collar workers (41%) were grouped together as presumably more positively oriented toward work than the blue-collar workers (23%). Fifth, an indication of the mother's attitude toward work was taken from the subject's reply to the question "What does your mother think is the purpose of college for a girl?" Replies were categorized according to whether or not the mother stressed preparation for an occupation.

The index of mother's work orientation was the sum of the scores on the five codes just described. Theoretically, scores could range from 5 to 10, the lowest score going to a subject's mother who was working at the time of the study, preferred to work, had a consistent history of being a working mother, had been employed in either professional or white-collar work, and stressed occupational preparation in her views on the purpose of college for the subject. The actual range was from 5 to 10, with a median of 7.5 and a decile range from 6 to 8.5.

For an index of *parent's views on the purpose of college,* responses were examined to the question "What does your father feel is the purpose of college for a girl?" and to the same question regarding the mother's opinion. Each response was coded according to whether the parent included vocational preparation among the purposes of college, which was the case for about two-thirds of the fathers and just over one-half of the mothers.

The *index of parents' attitudes toward importance of education for the daughter* was based on the subject's replies to questions about her parents' attitudes toward early marriage and about their views of the relative importance of education for boys and for girls. Theoretically scores on this index could range from 3 to 13, the lower score representing a subject whose father and mother both strongly oppose early marriage and who consider education more important for a girl than a boy. In the sample, scores ranged from 4 to 11 (no parents considered education more important for a girl than for a boy), with a median of 5 and a decile range from 5 to 8.

## RESULTS

The prevalence of orientation toward work among the subjects is indicated by the data displayed in Table 1. By five of the six criteria, 70% or more of the subjects were oriented toward future employment. For only a small minority of the subjects (14%), however, would there be any hesitation if there were a choice between marrying a particular man and continuing to work.

Intercorrelations among the six individual codes of work orientation are displayed in Table 2, along with part-

TABLE 1

WORK ORIENTATION AS EXHIBITED BY RESPONSES TO SEVERAL ISSUES AND AS ASSESSED BY INTERVIEWER

| Measure | Percentage work oriented | Percentage not work oriented |
|---|---|---|
| 1. Purpose of major curriculum | 77 | 23 |
| 2. Work in plans for near future | 88 | 12 |
| 3. Marriage and work | 70 | 30 |
| 4. Motherhood and work | 77 | 23 |
| 5. Conflict between husband and work | 14 | 86 |
| 6. Interviewer assessment of work motivation | 70 | 30 |

whole correlations between the individual codes and the index of work orientation.[4]

The intercorrelations among the indices are displayed in Table 3. The purpose of constructing these indices was to determine the relation between work orientation and five family background characteristics. In this sample, there was no significant correlation between work orientation and the following four family background characteristics: socioeconomic status (rho = .10), parents' educational level (rho = −.02), parents' views on purpose of college (rho = .01), and parents' attitudes toward importance of education for the daughter (rho = .13). There was a significant correlation between work orientation and the other family background characteristic, mother's work orientation: rho = .35, $t = 2.15$, $p < .05$, two-tailed test (Siegel, 1956, pp. 202–213).

## DISCUSSION

The descriptive data do not confirm the widespread observation that young college women are unrealistic about their futures. Their orientation is not discrepant from the likelihood that they will spend large segments of their lives in paid employment. It would be correct to say, however, on the basis of our impressions from the interviews, that their expectations of the future are vague and somewhat stereotyped. Like most college women today (cf. Newcomer, 1959), the subjects were secure in their expectations of marrying, and it would seem further that their views of their own futures do not extend very vividly beyond the period of involvement in childrearing.

Hindsight suggests that in place of using a single index of work orientation,

TABLE 2

INTERCORRELATIONS AMONG MEASURES OF SUBJECTS' WORK ORIENTATION

| Measure | Tetrachoric correlations 1 | 2 | 3 | 4 | 5 | 6 | Index of work orientation |
|---|---|---|---|---|---|---|---|
| 1. Purpose of major curriculum | — | .48 | .36 | −.07 | .10 | −.05 | $r_{pb} = .31$ |
| 2. Work in plans for near future | | — | .58 | .22 | .23 | .55 | $r_{pb} = .60$ |
| 3. Marriage and work | | | — | .53 | .40 | .45 | rho = .57 |
| 4. Motherhood and work | | | | — | .10 | .64 | $r_{pb} = .52$ |
| 5. Conflict between husband and work | | | | | — | .43 | $r_{pb} = .56$ |
| 6. Interviewer assessment of work motivation | | | | | | — | rho = .86 |

---

[4] The reader is reminded that point-biserial correlation coefficients ($r_{pb}$) tend to be lower than other familiar correlation coefficients; they are not estimates of $r$.

TABLE 3

INTERCORRELATIONS AMONG INDICES

(Spearman rank-correlation coefficients)

| Index | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. Subject's work orientation | — | .10 | −.02 | .35* | .01 | .13 |
| 2. Socioeconomic status | | — | (.80)* | .06 | −.09 | −.30 |
| 3. Parents' educational level | | | — | .12 | −.03 | −.25 |
| 4. Mother's work orientation | | | | — | (.34)* | −.01 |
| 5. Parents' views on purpose of college | | | | | — | .14 |
| 6. Parents' attitudes toward importance of education for the daughter | | | | | | — |

Note.—Correlations which are based in part on common elements are shown in parentheses.
* $p \leq .05$, two-tailed test.

it might have been better to distinguish two variables: *expectation* of employment and *aspiration* for achievement in employment. The impression garnered from the interviews is that most young women fully expect to work, but that for many the purpose is not so much to give service through a profession or to gain personal satisfaction from achievement as it is to conform through playing the feminine role now socially defined for them.

Of the five background variables examined, only the mother's work orientation was found to be correlated with work orientation of the subjects. Although it is hazardous to attach much confidence to a "significant" finding occurring in the context of a number of insignificant ones, this correlation is of interest, confirming the general view of the continuity of the generations through identification and role modeling within the family. However, one would have more confidence in this finding, and indeed in all findings which relate a characteristic of the subject to characteristics of her parents, if the information concerning the parents were obtained independently. Here we risk, in reverse, the same error which characterizes so much research on childrearing, in which parents are used as informants not only on their own characteristics but on those of their children as well.

In part, the generally negative result of the search for familial correlates may be attributable to the homogeneity of the subjects. They were much alike in their responses to the items comprising the index of work orientation (80% fell in a five-point range on an index with a theoretical range of 11 points) and were homogeneous in socioeconomic background and parents' educational level as well. Although the scores on the various indices were not so homogeneous as to render statistical analysis superfluous, they were sufficiently so to depress observed correlations. This may have operated in conjunction with the limited sample size to mask whatever relations may exist in the population among the variables.

## REFERENCES

HOLLINGSHEAD, A. B., & REDLICH, F. C. *Social class and mental illness.* New York: Wiley, 1958.

NEWCOMER, MABEL A. *A century of higher education for American women.* New York: Harper, 1959.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

(Received February 5, 1962)

*Journal of Educational Psychology*
1963, Vol. 54, No. 1, 38–44

# COMPARATIVE AND DEVELOPMENTAL STUDY OF THE LEARNING ABILITIES OF NEGRO AND WHITE CHILDREN UNDER FOUR CONDITIONS[1]

IRA J. SEMLER[2]
*Cedar Rapids Community School District, Iowa*

AND IRA ISCOE
*University of Texas*

Object pairs and picture pairs of these same objects in conceptually similar or dissimilar sets provided 4 experimental conditions for comparing paired-associate learning of 135 Negro Ss with 141 white Ss across age levels 5 through 9 years. Despite significantly lower Full Scale WISC IQs ($p < .001$), the analysis of variance showed that differences in paired-associate learning favoring white Ss at younger age levels disappeared by 9 years of age ($p < .05$) with no race difference when the age factor was disregarded. Race differences in IQ were greatest at 5 years, however, negligible correlations were obtained between learning and IQ for both races (.094 and .189). Associating object pairs in conceptually similar sets was easiest and picture pairs in dissimilar sets most difficult for both races ($p < .001$).

Measured differences between Negroes and whites continues to be a controversial topic in the psychological and educational literature. Garrett (1945a, 1945b, 1947) and more recently Shuey (1958) have pointed out the generally inferior performance of Negroes on a wide assortment of intelligence tests, both group and individual. Dreger and Miller (1960), in an extensive review of psychological differences between Negroes and whites, while not disagreeing completely with the genetic explanations of Garrett and Shuey have stressed strongly the importance of the psychocultural environment and the need for further cross-cultural research. There is little mention in this review of studies involving the learning abilities of Negroes and whites. This is not due to the reviewers' laxity since the typical Negro-white study is limited characteristically to reports of existing intellectual differences without regard to the effect of such differences on specific task performance.

The concern of the present research was to compare the abilities of Negro and white children on paired-associate learning tasks where previous measurement had determined that the white children had significantly higher IQs. However, since the relationship between measured intellectual status and learning ability is imperfectly understood (Sarason & Gladwin, 1958) and since paired-associate learning ability does not seem to be closely related to intelligence level (Berkson & Cantor, 1960; Eisman, 1958), it was hypothesized that no difference would be found between Negro and white children.

---

[1] This research was supported by a grant from the Human Genetics Foundation.

Portions of this paper were read at the Annual Convention, American Psychological Association, New York City, September 1961.

[2] Appreciation is expressed to administrative officials of the Austin Public School System for their cooperation. A special note of thanks is due C. L. Moore, Principal of L. L. Campbell School, and W. C. Wade, Principal of Becker School. Virgie Lee Washington of Greater Mt. Zion Nursery School was most helpful in all stages of the investigation.

## METHOD

### *Subjects*

All subjects were enrolled in the public schools and private kindergartens in Austin, Texas. At the time of the study the schools were still segregated as a result of neighborhood housing situations rather than school policy. An attempt was made to select schools so as to minimize socioeconomic differences. Despite these efforts and not unexpectedly for this region of the country, a significant difference in favor of the white children was obtained.

The ages of the children ranged from 5 through 9 years. The mean age for the whites was 7-2 and for the Negroes 7-4. The subjects were randomly selected by ages from lists furnished by the schools. Each child was first administered a Wechsler Intelligence Scale for Children (WISC) using trained examiners. After approximately a week, each child was administered the paired-associate learning task having been randomly assigned previously to one of the four experimental conditions. In all some 275 children were tested for this portion of the study. Table 1 provides comparative WISC Full Scale IQ data for the 275 subjects by sex and age level. The superiority of the whites, especially at the lowest age level, is to be discussed in detail later.

### *Selection of Stimulus Pairs*

Six pairs of stimuli were employed for the paired-associate learning task. Each stimulus pair was selected from a different conceptual classification involving items which were familiar to both groups and assumed to be informally learned early in the lives of all these subjects. Pretesting revealed that young children could sort the 12 stimulus objects into 6 pairs, based on functional utility or conceptual similarity, and could, in most cases, verbalize the basis for their pairings.

TABLE 1

MEAN WISC FULL SCALE IQs OF WHITE AND NEGRO CHILDREN BY SEX AT FIVE AGE LEVELS

| Subjects | Age in years | | | | | Total |
|---|---|---|---|---|---|---|
| | 5 | 6 | 7 | 8 | 9 | |
| White (*N* = 141) | | | | | | |
| Boys | | | | | | |
| *M* | 108.3 | 104.2 | 97.7 | 105.2 | 98.5 | |
| *N* | 19 | 22 | 12 | 15 | 12 | 80 |
| Girls | | | | | | |
| *M* | 116.5 | 104.9 | 97.2 | 105.8 | 94.2 | |
| *N* | 7 | 15 | 15 | 12 | 12 | 61 |
| Negro (*N* = 134) | | | | | | |
| Boys | | | | | | |
| *M* | 83.4 | 92.6 | 90.6 | 98.3 | 89.5 | |
| *N* | 13 | 14 | 12 | 20 | 12 | 71 |
| Girls | | | | | | |
| *M* | 82.5 | 103.0 | 90.8 | 96.5 | 85.0 | |
| *N* | 13 | 12 | 12 | 14 | 12 | 63 |

The six pairs employed were:

| Stimulus | Response |
|---|---|
| BANANA | ORANGE |
| TOOTHBRUSH | COMB |
| PIPE | CIGAR |
| DOLL | TOY TRUCK |
| CUP | BOWL |
| GLOVE | SHOE |

The assumption was that the above pairs possess high associative value by virtue of conceptual or functional similarity. In order to provide item pairs which would require new learnings (associations) these Similar pairs were rearranged as follows:

| Stimulus | Response |
|---|---|
| BANANA | SHOE |
| TOOTHBRUSH | ORANGE |
| PIPE | COMB |
| DOLL | CIGAR |
| CUP | TOY TRUCK |
| GLOVE | BOWL |

The assumption underlying the resulting Dissimilar condition was that the associative values in terms of categorical or functional similarity are much lower than in the case of the Similar condition. This assumption is supported by the findings of Underwood and Richardson (1956) with regard to verbal associations and concept formation.

In order to assess the effect of the availability of cue on the learning of the stimulus pairs employed, one set was composed of actual objects (Concrete) while the other consisted of high quality colored 3¼ × 4¼ inch photographs (Abstract) of these same objects mounted on 5 × 7 inch neutral colored cardboard. Therefore, the labels Concrete and Abstract refer to operationally defined stimulus characteristics. Combinations of the two sets of stimulus pairs (Similar versus Dissimilar) and two degrees of cue presence (Concrete versus Abstract) were used to establish the four experimental conditions—i.e., Similar-Concrete, Similar-Abstract, Dissimilar-Concrete, and Dissimilar-Abstract.

### *Design and Procedure*

Stimulus pairs were presented manually for all conditions. The stimulus object or picture was shown for approximately 5 seconds followed by the response object or picture for approximately 3 seconds regardless of the correctness of the subject's response. The intertrial interval was approximately 5 seconds. Each subject was run for 12 blocks of 6 pair presentations with a total of 72 pair presentations in all. Serial learning effects were avoided by randomizing order of pair presentations across trial blocks. An observer recorded the subject's response on each trial and the total number of correct anticipations in the 6 pair presentations for 12 blocks was used in the analysis of variance. The 275 subjects who had been administered the WISC previously were randomly assigned to one of the four experimental conditions and given the paired-associate learning task appropriate to their assignment. From this subject pool 31 Negro children and 31 white children were randomly selected from each of the four conditions providing 124 white and 124 Negro subjects for a race by condition analysis. Since analyzing data from such a design furnishes no age or race by age information a three-factor repeated measurement design was employed involving two races, four conditions, and five age levels. In order to equalize the number of subjects in each cell of this design, a proportional analysis using 120 (60 white subjects and 60 Negro subjects) was carried out. In addition, a race by age analysis of measured intelligence for these same 120 subjects was made in order to compare acquisition rate and intellectual level. A correlation analysis between intelligence and total learning score for each race was also computed for the original 275 subjects.

## RESULTS

The analysis of variance for the 124 white and 124 Negro subjects under the four conditions yielded an $F$ of 81.73 which, with 3 and 240 degrees of freedom, is significant at the .001 level. Nonsignificant $F$s for race and the race by condition interaction were obtained. It is apparent that, while the conditions yielded significantly different rates of learning, the two races did not differ in this respect nor was any one condition significantly more difficult for either race.

The proportional analysis of variance for the 120 subjects which supports the previous analysis in greater detail is summarized in Table 2.

Again, the four experimental conditions yielded highly significant differences. The significant age effect ($p < .001$) and the race by age interaction ($p < .05$) are graphically illustrated by Figures 1 and 2.

In terms of age trends, the learning rate of white children was superior to Negro children under the Similar condition at the 5- and 6-year-old level (Figure 1) for both Concrete and Abstract paired-associate presentations. In addition, differences may be noted between whites and Negroes under the Dissimilar-Concrete and the Dissimilar-Abstract conditions. However, race differences in learning ability decreased steadily across the age levels studied and were not significant in the summary analysis. This is illustrated by comparing the learning curves of Figure 1 for the combined 5-

TABLE 2

SUMMARY OF THE PROPORTIONAL ANALYSIS OF VARIANCE OF PAIRED-ASSOCIATE LEARNING OF WHITE AND NEGRO CHILDREN BY AGE UNDER FOUR CONDITIONS

| Source | $df$ | $MS$ | $F$ |
|---|---|---|---|
| Between subjects | 119 | | |
| Race | 1 | 6.137 | .770 |
| Condition | 3 | 506.858 | 63.628**** |
| Age | 4 | 44.445 | 5.579**** |
| Race × Condition | 3 | 5.815 | .730 |
| Race × Age | 4 | 20.768 | 2.607* |
| Condition × Age | 12 | 6.633 | .833 |
| Race × Condition × Age | 12 | 7.760 | .974 |
| Error (b) | 80 | 7.966 | |
| Within subjects | 1320 | | |
| Trials | 11 | 166.916 | 200.139**** |
| Trials × Race | 11 | .849 | 1.018 |
| Trials × Condition | 33 | 7.001 | 8.394**** |
| Trials × Age | 44 | .865 | 1.037 |
| Trials × Race × Condition | 33 | .885 | 1.061 |
| Trials × Race × Age | 44 | .836 | 1.002 |
| Trials × Condition × Age | 132 | .525 | .629 |
| Trials × Race × Condition × Age | 132 | 1.031 | 1.236 |
| Error (w) | 880 | .834 | |
| Total | 1439 | | |

* $p < .05$.
**** $p < .001$.

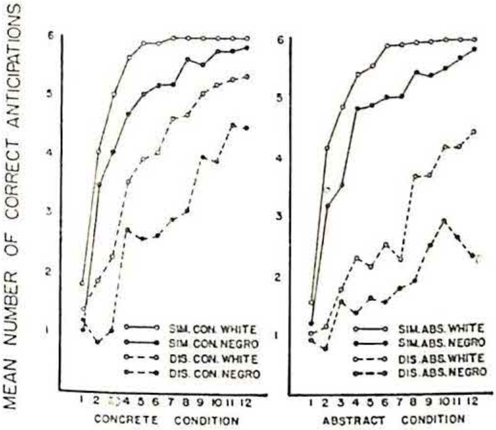


FIG. 1. Trials in blocks of six 5- and 6-year-olds.

and 6-year-olds with the curves in Figure 2 for the combined 8- and 9-year-olds where learning ability differences have disappeared. A further analysis of race differences at the five age levels in terms of mean difference in total learning score is reported in Table 3 which shows in greater detail the developmental trend differences indicated by the significant race by age interaction reported in Table 2.

Figures 1 and 2 also show that learning was most rapid for both white and Negro 5- and 6-year-old children under the Similar-Concrete condition. This is the condition wherein the pairs to be associated were categorically similar and cue presence was maximized. The rate of acquisition under the Similar-Abstract condition was almost as rapid for both groups as under the Similar-Concrete condition. Cue reduction, regardless of race, does not exert a differential effect on the learning of conceptually similar pairs. However, the effect of pair dissimilarity was to lower learning rates. This was more pronounced under the Dissimilar-Abstract condition than under the Dissimilar-Concrete condition and was true for both races. It appears, then, that cue reduction has a significant effect on acquisition rate when the pairs to be associated involved new learning as was the case under the Dissimilar condition.

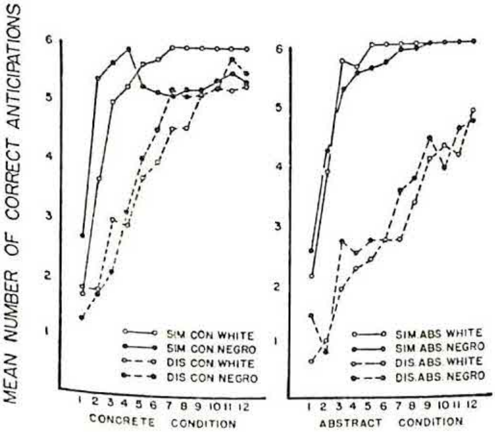


FIG. 2. Trials in blocks of six 8- and 9-year-olds.

TABLE 3

MEAN OF TOTAL LEARNING SCORES AT FIVE AGE LEVELS

| Age level | | White | Negro | $t$ |
|---|---|---|---|---|
| 5 | $M$ | 48.92 | 38.25 | 10.67** |
| | $N$ | 12 | 12 | |
| 6 | $M$ | 56.17 | 50.08 | 3.73** |
| | $N$ | 12 | 12 | |
| 7 | $M$ | 52.67 | 56.42 | −2.30* |
| | $N$ | 12 | 12 | |
| 8 | $M$ | 54.16 | 56.42 | −1.39 |
| | $N$ | 12 | 12 | |
| 9 | $M$ | 50.33 | 53.25 | −1.79 |
| | $N$ | 12 | 12 | |

Note.—Based on proportional analysis of variance.
* $p < .05$.
** $p < .01$.

In general, most of the between-subjects and within-subjects interactions were nonsignificant leaving "pure" main effects to be studied. The significant trials by condition interaction is clearly seen in a comparison of the learning

TABLE 4

CORRELATION OF WISC IQS WITH TOTAL LEARNING SCORE

| Race | Verbal | Performance | Full Scale |
|---|---|---|---|
| White ($N = 141$) | .033 | .127 | .094 |
| Negro ($N = 134$) | .117 | .204* | .189* |

* Reliably different from 0 at the .05 level of confidence.

curves in Figures 1 and 2 where acquisition rate differs depending upon the experimental treatment under consideration.

In order to determine the relationship between total learning score and intellectual level a comprehensive correlation analysis was carried out for both whites and Negroes. Part of these data is reported in Table 4 which indicates low but reliable correlations between WISC Performance and Full Scale IQs with total learning score for Negroes only. Since very low but reliable correlations were obtained, a further breakdown of Verbal, Performance, and Full Scale WISC IQs by race and age is reported in Table 5. It should be noted that the data in Table 5 were based on the 120 subjects used in the proportional analysis of variance of Table 2. Further, an analysis of variance of WISC Full Scale IQs for these same 120 subjects is reported in summary form in Table 6. This indicates a highly significant difference in WISC Full Scale IQ in favor of white subjects and significant differences across the age levels studied. However, the differences were more pronounced at the 5-year-old level which may partially account for the race by age interaction in view of the low (but reliable) correlation obtained between total learning and intelligence level. It is doubtful, however, if this is the main factor in the race by age interaction.

TABLE 5

MEAN VERBAL, PERFORMANCE, AND FULL SCALE WISC IQS OF THE 120 SUBJECTS AT FIVE AGE LEVELS USED IN THE PROPORTIONAL ANALYSIS OF VARIANCE

| WISC IQ | Age in years | | | | |
|---|---|---|---|---|---|
| | 5 | 6 | 7 | 8 | 9 |
| White ($N = 60$) | | | | | |
| Verbal | 110.75 | 104.75 | 102.67 | 102.00 | 91.33 |
| Performance | 116.33 | 110.83 | 98.08 | 109.58 | 93.25 |
| Full Scale | 114.67 | 108.33 | 100.58 | 106.08 | 91.42 |
| $N$ | 12 | 12 | 12 | 12 | 12 |
| Negro ($N = 60$) | | | | | |
| Verbal | 87.75 | 105.50 | 92.50 | 99.42 | 90.17 |
| Performance | 86.92 | 97.25 | 89.83 | 100.00 | 85.50 |
| Full Scale | 86.00 | 101.83 | 90.50 | 99.50 | 86.75 |
| $N$ | 12 | 12 | 12 | 12 | 12 |

TABLE 6

SUMMARY ANALYSIS OF VARIANCE OF FULL SCALE WISC IQS FOR THE 120 SUBJECTS USED IN THE PROPORTIONAL ANALYSIS OF VARIANCE IN TABLE 2

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Race | 1 | 3830.700 | 27.472**** |
| Age | 4 | 975.138 | 6.993**** |
| Race × Age | 4 | 588.554 | 4.221*** |
| Cells | 9 | | |
| Within subgroups (w) | 110 | 139.436 | |
| Total | 119 | | |

*** $p < .005$.
**** $p < .001$.

## DISCUSSION

The lack of previous studies of a similar nature produces some difficulty in the evaluation of the findings of this study. It is not the purpose of this discussion to emphasize the race differences obtained in measured intelligence. The data indicate the limitation of measured intelligence in predicting performance under conditions where new learning is to take place. In the case of this study, the procedure and design suggest a variety of fruitful ways of studying factors which may have a significant effect on the acquisition of new learning. This is an important consideration in studies between races or between groups. The results provide developmental support, however, for a hypothesis of no difference in learning ability between the groups studied under the four experimental conditions.

Specifically, the results suggest that young Negro children have more difficulty than young white children in learning new associations, especially when cues are reduced by using photographs rather than actual objects. The greatest discrepancy in Full Scale IQ (about 28 points) existed at the 5-year-old level. However, in view of the decreasing difference between races in learning ability across age levels, and in spite of continuing significant differences in intelligence, and the low correlations between IQ and total learning score, it seems improbable that intellectual differences alone could explain the superior learning of the young white subjects. Some consideration is due, therefore, for environmental factors.

All of the 5-year-old and some of the 6-year-old subjects were obtained from nursery schools. Public school kindergartens per se did not exist and a child must be age 6 by September 1 of the school year to enter first grade. As far as nursery school attendance goes, the great majority of Negro children attend because the mother works. The Negro 5-year-old children were taken from a setting run by a church with over 120 children ranging in ages from 3 through 6. They were cared for all day in less than adequate circumstances. In contrast, the white children were not for the most part children of working mothers. Whatever their parents' reasons for sending them to nursery school, the facilities in terms of physical plant and staff were excellent. The children received some informal school training such as learning to write numbers and letters. The implication here is that the environment of the 5-year-old Negro children was not as stimulating towards new learning situations as that of the whites. There is also the added implication that, in terms of socioeconomic difference, the Negro children at the 5-year-old level were much lower than the whites. In terms of our overall sample of Negro and whites, the Negroes were also of lower socioeconomic status. This is, of course, a problem in comparing racial groups. It is virtually impossible to equate Negroes and whites on the basis of socioeconomic status. Standard of living, steadiness of employment, and number of persons in the family working all complicate the estimate of economic status.

Some comment is due also concerning the older Negro children and their public school facilities. They came from a school which had a very adequate physical plant, a good lunch program, and a dedicated, well-trained group of teachers. This school was certainly the equal of that of our older white children. Whether these factors really contributed to the gradual erasing of differences in learning ability with increasing age cannot be answered unequivocally, but they should at least be pointed out. A combination of environmental factors may well have contributed to the comparatively poor overall learning of the younger Negro children. At any rate, there is little to be gained by falling back on a genetic explanation inasmuch as these too cannot be separated from environmental conditions. What is important to note is the developmental support obtained in this study for no differences in overall learning with increasing age of our subjects.

There is a certain timeliness in the results of the present research. As new educational opportunities are opened for the Negro child in the United States, educators will be challenged to find methods whereby the optimum conditions for learning can be found and applied. The advent of integration has already produced a great many problems concerned with the learning and motivation of minority groups, especially Negroes. Our findings of no overall race differences in learning ability should not be minimized. They suggest that educators should exercise great caution in inferring learning ability from measured intellectual level alone. The developmental study of specific learning abilities of Negroes and whites could provide a more effective basis for the psychological and educational decisions that inevitably will have to be made by those who bear the responsibility of discovering and implementing the conditions for the maximum development of human potential. One study does not in itself prove anything definitive, especially in an area that has been replete with charges and counter-charges in the past. Further research, including possible replication of the present study, is indicated.

## REFERENCES

BERKSON, G., & CANTOR, G. N. A study of mediation in mentally retarded and normal school children. *J. educ. Psychol.*, 1960, **51**, 82–86.

DREGER, R. M., & MILLER, K. S. Comparative psychological studies of Negroes and whites in the United States. *Psychol. Bull.*, 1960, **57**, 361–402.

EISMAN, BERNICE S. Paired associate learning, generalization and retention as a function of intelligence. *Amer. J. ment. Defic.*, 1958, **63**, 481–489.

GARRETT, H. E. A note on the intelligence scores of Negroes and whites in 1918. *J. abnorm. soc. Psychol.*, 1945, **40**, 344–346. (a)

GARRETT, H. E. Psychological differences as among races. *Science*, 1945, **101**, 16–17. (b)

GARRETT, H. E. Negro-white differences in mental ability in the United States. *Scient. Mon.*, 1947, **65**, 329–333.

SARASON, S., & GLADWIN, T. Psychological and cultural problems in mental subnormality: A review of the research. *Genet. Psychol. Monogr.*, 1958, **57**, 7–269.

SHUEY, AUDREY M. *The testing of Negro intelligence.* Lynchburg: Randolph Macon Women's College, 1958.

UNDERWOOD, B. J., & RICHARDSON, J. Some verbal materials for the study of concept formation. *Psychol. Bull.*, 1956, **53**, 84–95.

(Received July 10, 1962)

*Journal of Educational Psychology*
1963, Vol. 54, No. 1, 45-51

# STUDENTS' THOUGHTS DURING CLASS: A CRITERION FOR EDUCATIONAL RESEARCH[1]

LAURENCE SIEGEL, LILA CORKLAND SIEGEL, PATRICK J. CAPRETTA,
REGINALD L. JONES, AND HOWARD BERKOWITZ
*Miami University, Oxford, Ohio*

Relevance of student thinking during class presentations is proposed as a relatively sensitive and uncontaminated criterion for educational research. Reports of thinking were evoked by modifying the stimulated recall technique using video tape and group administration. These reports were independently classified along a continuum of "thinking relevance" by pairs of judges, and the resultant classifications were transformed to numerical weights determined by a paired-comparisons analysis. Reliability and validity of the stimulated recall procedure were evaluated across lectures ($n = 3$) and "critical stops" within lectures ($n = 12$). The median interrater reliability for critical stops was .58. Validity against an independent measure of student learning was .61. The use of at least 3 trials reflecting variations in student mood and interest is recommended.

Of all the problems besetting psychoeducational research, the criterion problem is certainly one of gravest importance. The search for appropriate research criteria has often had the beneficial effect of focusing the attention of both investigator and teacher upon desired instructional outcomes. These outcomes are then reflected in the structure of course examinations used to assess what students have learned. Occasionally, course examinations are supplemented with instruments designed to measure such less tangible or more elegant objectives as attitudinal change and critical thinking. Regardless of their nature, the overwhelming emphasis in developing criteria for educational research is upon instruments administered at or very near the completion of the course.

Such postcourse criterion measures may suffer considerable contamination in higher educational settings. The investigator typically wishes to evaluate the effectiveness of some instructional procedure from student performance on the final examination. Realistically, however, final examination performance reflects a variety of variables other than the ones under study. To cite an extreme instance of such criterion contamination, it is quite possible for students to earn high final examination scores without ever attending a class meeting. This is so because they have access to sources of information other than face-to-face contact with the instructor: textbooks, and comprehensive notes taken by a friend who *does* attend class are but two such sources.

Furthermore, although a final examination may be carefully structured in accord with instructional objectives, it may lack sensitivity as a research criterion. This would be the case, for example, when the effects of the independent variable are cumulative and gather strength throughout the semester. Since postcourse examination scores typically reflect learning throughout the entire course when the independent variable effects have been both weak and strong, such examinations have a tendency to mask cumulative effects (Siegel, 1960).

---

[1] The research reported herein was a byproduct of a larger study of the Instructional Gestalt supported by a grant from the Office of Education, United States Department of Health, Education, and Welfare.

These two defects of postcourse examinations as research criteria, contamination and lack of sensitivity, suggest the importance of an entirely different kind of criterion. Certain types of investigations could be enhanced by utilizing an immediate rather than a delayed criterion measure, reflecting the impact of independent variables in situ and thereby free from contamination by extraneous factors.

One such criterion, students' in-class thought processes, has been suggested and used by Bloom (1953). He tabulated categories of student thinking during class periods in an attempt to make a diagnostic assessment of the relative strengths and weaknesses of lecture versus discussion presentations of subject matter. The technique for evoking records of student thinking is termed *stimulated recall.* The entire class period was recorded on audio tape. These tapes were subsequently played during individual interviews with students within 24 hours after the original class period. The tape was stopped at "critical points" and the interviewee was asked to attempt to relive the original experience and recall what he was thinking during the stimulus lecture or discussion.

The stimulated recall technique described above was suitable for Bloom's purpose but lacked the refinements required of criterion instruments. It yielded percentage summaries of student thoughts assigned in terms of a classificatory scheme rather than scores along some kind of continuum of student thinking. The present paper is concerned with a modification of the general technique of stimulated recall and summarizes evaluative data concerning its usefulness as a research criterion.

### *Problem*

Stimulating recall by audio tape with delayed playback to evoke a record of thinking during the original classroom presentation suffers from at least three major difficulties. First, the period of delay (up to a maximum of 24 hours) between the initial presentation and playback may increase the likelihood that respondents will forget or suppress the thoughts they experienced during the original presentation. Second, audio tape reproduces only a portion of the original classroom experience. To the extent that cues in addition to auditory ones could be introduced to stimulate recall, the resultant memory ought to be more complete. Finally, the technique of collecting data from individual student interviews is obviously laborious.

In order to determine whether these three difficulties could be circumvented, we employed a technique whereby (*a*) stimulated recall trials were conducted almost immediately following the original presentation, (*b*) the original presentation was recorded on video tape and students observed the playback from the same seats in the same classroom environment as the original presentation, (*c*) data were simultaneously collected in written form from the entire group of students enrolled in the course.

The basic problem was to evaluate the utility of stimulated recall trials thus conducted for providing a meaningful criterion of student thinking in situ. This evaluation required a series of investigations to (*a*) design a scoring system reflecting relevance of thinking, (*b*) ascertain the reliability with which raters could convert these written records of thinking to the numerical scale of relevance, (*c*) determine the consistency of students' stimulated recall scores across trials encompassing different subject matter presentations on different days, and (*d*) estimate the validity of the scores derived from the stimulated recall technique.

## METHOD

### *Subjects*

These investigations were conducted in four viewing sections of the televised course Foundations of Human Behavior offered for credit to students on the Miami University campus. These sections, each consisting of approximately 60 students, met for two weekly 50-minute TV presentations in a conventional classroom equipped with TV receivers. Virtually all of these students were freshmen enrolled in various teacher-training curricula.

### *Procedure*

Stimulated recall trials were not attempted until after mid-semester. This study was postponed until the second half of the semester to enable students to become used to the routine of attending a televised class. It was feared, were this not done, that some students might use stimulated recall trials merely as an outlet for expressing personal feelings about TV instruction.

The procedure for conducting a stimulated recall trial was as follows.

The stimulus lecture (20 minutes in length) was recorded on video tape several days in advance of the scheduled trial. This taped lecture did not differ in any way from the usual lecture except for its abbreviated length. During the taping, the experimenter observed from the studio and noted the position on the tape of possible "critical points." A critical point was grossly identified as any point in the lecture that would be likely to evoke some kind of student thinking. Operationally, at such points, the lecturer generally asked a question, defined a term, or attempted to synthesize and relate concepts. Finally, three to five critical points were selected from each lecture: the first of these 4–5 minutes after the lecture began, and the succeeding ones spaced at 3–5 minute intervals.

On the day of the stimulated recall trial, the taped lecture was presented on TV. Students were given no indication that the lecture was taped rather than live, that it was to run for only 20 minutes, or that anything unusual would occur during the class period.

The stimulus lecture terminated with the instructor introducing the experimenter to the class as a researcher. The experimenter asked the proctor who had just come into the receiving room to administer the tests with which he had been provided. One of these tests was designed to validate the stimulated recall procedure and is discussed subsequently in more detail.

Upon completion of the testing (approximately 20 minutes), the experimenter reappeared on camera to prepare the students for the stimulated recall trial. He explained that the lecture just seen had been recorded on video tape; it would be replayed from the beginning; students were to attempt to adopt the "frame of mind" they had at the beginning of the class period and "relive" the first 20 minutes of the hour; the tape would be interrupted at several critical points; at each of these stops the respondent was to write on the sheet provided him "whatever he was thinking at that point during the original presentation." Assurances were given concerning the confidential nature of the replies.

Following this introduction, the tape was replayed from the beginning, hopefully taking the viewers back about 45 minutes in time. It was stopped at the first critical point by prearrangement with the studio staff, and the experimenter repeated the instruction to "write down what you were thinking at this point during the original presentation." A time lapse of 2 minutes was allowed for students to write their thoughts, and the tape was started from the preceding stopping place. This procedure was followed for all stops within a trial.

Three such trials were conducted with this group of students. The first trial involved a portion of a lecture on visual defects and had three stops; the second stimulus lecture concerned fundamental concepts in psychological measurement and had four stops; the third lecture concerned projective techniques and had five stops. Thus for each subject we obtained his statement of thoughts in situ on 12 occasions, sampling three lectures spaced over approximately a 5-week period.

The validation test administered in one viewing section between the end of the stimulus lecture and the beginning of the stimulated recall playback was mentioned above briefly. This achievement test contained two kinds of items: "critical" items measuring knowledges conveyed at or immediately preceding each scheduled stimulated recall stop; and "control" items measuring knowledges conveyed within the chapter of the textbook from which the lecture was drawn but not discussed or implied in the lecture itself. Each of these tests contained 2 critical items for each stimulated recall stop and between 10 and

12 control items. Across the three trials, we obtained scores for each subject on 24 critical items and 34 control items.

The score on the control items reflected knowledges acquired in these three topical areas from extra-class sources. The score on the critical items reflected the influence of these extra-class sources of information in combination with exposure to the lecture. Hence, as a test of validity of stimulated recall, the total stimulated recall scores across the three trials (indicating relevance of thinking) were correlated with the critical item subscores partialing out the control item subscores. This partial correlation reflects the validity of students' reports about their thinking against a criterion of acquisition corrected for knowledges acquired from sources other than the class lecture.

## RESULTS

The findings are most conveniently summarized under four headings: Scoring for Relevance of Thinking, Interrater Reliability, Validity, and Intratrial and Intertrial Consistency.

### *Scoring for Relevance of Thinking*

Bloom (1953) suggested a system for coding (but not for weighting) thoughts evoked by his stimulated recall procedure. Our first step in developing a scoring system was to code the data using Bloom's categories. This coding was accomplished independently after a minimum amount of preliminary training by statistical clerks serving as judges. Two judges coded each stop.

The coding categories with illustrations of each are reproduced below. These are identical to Bloom's categories with the exception of Classification IIf, added to handle a number of responses in our study.

I. Irrelevant thoughts
   a. Thoughts about persons, objects and events not in the classroom environment ("I was wondering if John would ask me to go to the dance on Saturday.")
   b. Tangential thoughts about words and phrases used in the lecture ("Myopia sure is a funny sounding word." "Mr. X sure says 'uh' a lot while he lectures.")

II. Relevant thoughts
   a. Passive thoughts about the subject ("I was just taking notes." "Just listening.")
   b. Thoughts evidencing simple comprehension of the subject ("Somebody with tunnel vision would have to move his eyes around to see things at the sides.")
   c. Thoughts involving attempts to apply and utilize the subject matter ("I am nearsighted and was thinking about becoming farsighted later in life and probably having to wear bifocals.")
   d. Thoughts involving attempts to find solutions to problems or synthesize the subject ("He [instructor] said that much special training is required to interpret projectives correctly. I wonder how practical they are for routine use. Can't they be put in an objectively scorable form?")
   e. Thoughts involving evaluation of the meaningfulness and accuracy of the lecture ("I don't believe you can tell very much from a figure drawing. Talent would enter into what you draw.")
   f. Thoughts wherein questions beyond the presentation are asked ("Is it unusual for a person to make more than one response to either a part of a blot or the whole blot?")

In using this system of coding, each judge read the reported thought and decided first whether subject's thinking was relevant or irrelevant. Then he assigned the appropriate code designation indicating the category of relevant or irrelevant thinking into which he judged it to fit. Whenever a subject reported both relevant and irrelevant thoughts at a given stop, the judge always assigned the code designation only for the relevant thought.

Four experimenters independently rated the eight coding classifications using the paired-comparisons technique. Every classification was compared with every other classification and a judgment made about which represented the "more relevant" thinking. Similar paired-comparison judgments were made by seven persons who had considerable prior experience in actually coding stimulated recall responses.

Summaries of these two sets of paired-comparison ratings agreed closely. The code categories were distributed along a continuum of "thinking relevance" on the basis of these ratings in accord with the following numerical weights:

Irrelevant thoughts (Ia and Ib) = 0
Relevant thoughts
IIa, IIb = 2
IIc, IIe, IIf = 4
IId = 5

These weights were applied to the coded thought contents and summed across stops for each trial to yield "trial scores," and across the three trials to yield total scores. All analyses of stimulated recall data subsequently described involved the utilization of these weights and scores.

### *Interrater Reliability*

Each stimulated recall response was coded independently by two judges. Correlations between the numerical conversions for each pair of judgments are summarized for 225 subjects in Table 1.

Interrater reliabilities for the 12 stops ranged between .47 and .76 with a median value of .58. When total trial scores were considered, interrater agreement ranged from a low of .65 on Trial 3 to a high of .77 on Trial 1. Since our raters received relatively little preliminary training, it is likely that these coefficients are minimum estimates of the interrater reliability that may be anticipated in scoring stimulated recall data.

TABLE 1

INTERRATER RELIABILITIES

| Trial | Stop | *r* | |
|---|---|---|---|
| I | 1 | .69 | |
| | 2 | .76 | |
| | 3 | .70 | |
| Total score | | | .77 |
| II | 1 | .57 | |
| | 2 | .49 | |
| | 3 | .59 | |
| | 4 | .64 | |
| Total score | | | .70 |
| III | 1 | .62 | |
| | 2 | .55 | |
| | 3 | .47 | |
| | 4 | .56 | |
| | 5 | .55 | |
| Total score | | | .65 |

Discrepancies between the weights assigned to given responses by pairs of judges were arbitrated by experimenters prior to conducting the validity and consistency studies described in the following sections.

### *Validity*

The crux of the utility of stimulated recall as a criterion of student thinking is its validity. Do students' reports about what they were thinking at various points in the stimulus lecture correlate with an independent measure of what they have learned at these points *solely* as a result of the lecture? Presumably, students indicating that they were inattentive to the presentation ought to have learned very little from it. Conversely, students reporting highly relevant thinking should be expected to show a parallel gain in knowledge from the stimulus lecture.

As described earlier, the design for validation involved administration of an achievement test consisting of subsets of critical items and control items in one of the viewing sections ($N=47$). Performance on the control items was assumed to reflect information in the general area of the lecture topic acquired from sources other than the lecture presentation itself.

The correlation matrix for total scores (across the three trials) on critical items, control items, and stimulated recall is shown in Table 2. The correlation between stimulated recall and critical item scores partialing out the control item scores was .61. It is evident that stimulated recall scores do, in fact, correlate with a measure of achievement reflecting knowledges gained from the lecture itself. The higher the level of thinking relevance displayed by a student during stimulated recall, the greater is the likelihood that he will learn the knowledges conveyed during the lecture.

*Intratrial and Intertrial Consistency.*

The factors affecting the internal consistency of stimulated recall responses parallel those affecting the internal consistency of achievement test item responses. However, one aspect of stimulated recall is uniquely important: the character of student thinking may be presumed to be particularly sensitive to fluctuations of interest during a given stimulus lecture and in different lectures.

TABLE 2

INTERCORRELATIONS BETWEEN SCORES ON CRITICAL ACHIEVEMENT ITEMS, CONTROL ACHIEVEMENT ITEMS, AND STIMULATED RECALL

| Item | Critical items | Control items | Stimulated recall |
|---|---|---|---|
| Critical | X | .28 | .59 |
| Control | | X | .02 |

Note.—$N = 47$.

The present design sampled thinking during three different lectures and at three to five points within each lecture. The intercorrelations reflecting intratrial and intertrial consistency of thinking are exhibited in Table 3.

Although the means and standard deviations of relevance scores are relatively stable, the intratrial and intertrial

TABLE 3

INTRATRIAL AND INTERTRIAL CONSISTENCY OF STIMULATED RECALL SCORES

| Trial | Trial I | | | | Trial II | | | | | Trial III | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | Total | 1 | 2 | 3 | 4 | Total | 1 | 2 | 3 | 4 | 5 | Total |
| *M* | 1.73 | 1.88 | 1.58 | 5.20 | 1.39 | 1.76 | 1.59 | 1.63 | 6.44 | 1.31 | 1.54 | 1.52 | 1.41 | 1.47 | 7.33 |
| *SD* | 1.34 | 1.35 | 1.39 | 2.97 | 1.27 | 1.28 | 1.35 | 1.42 | 3.32 | 1.17 | 1.07 | 1.09 | 1.16 | 1.00 | 3.22 |
| I | | | | | | | | | | | | | | | |
| Stop 1 | X | .35 | .22 | | | | | | | | | | | | |
| Stop 2 | | X | .34 | | | | | | | | | | | | |
| Stop 3 | | | X | | | | | | | | | | | | |
| Total score | | | | | | | | | .25 | | | | | | .29 |
| II | | | | | | | | | | | | | | | |
| Stop 1 | | | | | X | .27 | .28 | .15 | | | | | | | |
| Stop 2 | | | | | | X | .23 | .13 | | | | | | | |
| Stop 3 | | | | | | | X | .13 | | | | | | | |
| Stop 4 | | | | | | | | X | | | | | | | |
| Total score | | | | | | | | | | | | | | | .29 |
| III | | | | | | | | | | | | | | | |
| Stop 1 | | | | | | | | | | X | .14 | .13 | .13 | .08 | |
| Stop 2 | | | | | | | | | | | X | .28 | .16 | .19 | |
| Stop 3 | | | | | | | | | | | | X | .30 | .20 | |
| Stop 4 | | | | | | | | | | | | | X | .35 | |
| Stop 5 | | | | | | | | | | | | | | X | |

Note.—$N = 225$.

correlations tend to be fairly low. Hence, there is no strong evidence for the existence of characteristic patterns or "styles" of thinking.

The practical implication of this finding for the use of stimulated recall to provide a research criterion involves the sampling problem. How many trials involving different stimulus lectures are required to obtain a reasonably stable measure of student thinking? This issue may be viewed from a prediction standpoint. From this perspective, the question becomes: "How many trials [t] are required to reach the point where the multiple correlation $R_{t \cdot t-1}$ is essentially identical to $R_{t \cdot t+1}$?"

Our data were not sufficient to answer this problem definitively. The correlation between Trial 1 and Trial 2 scores was .25. The multiple correlation using Trials 1 and 2 as joint predictors of Trial 3 was .41. We have no way of knowing what the multiple correlation using Trials 1, 2, and 3 in combination to predict a fourth trial would have been. However, we infer from the relatively substantial validity coefficient obtained for the three-trial total, that this multiple correlation would have been substantially higher than .41 if the data had been available.

## DISCUSSION AND CONCLUSIONS

Student thinking evoked by stimulus lectures and assessed for relevance by means of a stimulated recall technique holds promise as a research criterion. Although the resultant data are subjectively scored, interrater reliabilities were reasonably high. Furthermore, the resultant scores for thinking relevance had demonstrated validity against an objective measure of knowledges acquired from the stimulus lecture.

This research leads to the conclusion that student thinking sampled at various points within a lecture and across lectures is relatively inconsistent with respect to relevance. The implication of this finding for use of the stimulated recall technique to provide criterion data is that it is desirable to sample across a variety of topics. Although the optimal number of required trials is not known, the present investigation supports the use of at least three trials to reflect variations in student mood and interest.

## REFERENCES

BLOOM, B. S. Thought processes in lectures and discussions. *J. gen. Educ.*, 1953, **7**, 160–169.

SIEGEL, L. The instructional gestalt: A conceptual framework. *Teachers Coll. Record*, 1960, **62**, 202–213.

(Received July 10, 1962)

*Journal of Educational Psychology*
1963, Vol. 54, No. 1, 52–56

# NOTE ON SOCIAL DESIRABILITY RESPONSE STYLE AND LEARNING

LAWRENCE J. STRICKER

*Educational Testing Service, Princeton, New Jersey*

In Brown's (1960) study of the learning of personality items which were paired with nonsense syllables, a major finding was that social desirability response style was related to learning measures for women but not men. A reanalysis of these results found that social desirability response style was related to delayed recall for both men and women, and confidence and immediate recall for women only. It was not related to trials to mastery, learning time, and blockings. It was concluded that this study partially supports Edwards' social desirability response style reinterpretation of studies of anxiety and learning, but that many previous studies do not.

The relationship between social desirability response-style and learning performance has been investigated by Brown (1960) in a study of paired-associate learning. Men and women who were classified as high or low in social desirability response style on the basis of their scores on Edwards' (1957) Social Desirability (*SD*) scale learned 16 statements from the Edwards Personal Preference Schedule (EPPS, Edwards, 1953) which were paired with nonsense syllables. The statements represented two levels of social desirability and two levels of descriptiveness (the latter variable was derived from each subject's classification of the statements as "like me" or "not like me").

Brown reported the results for four dependent variables: (*a*) confidence[1]—subject's estimate, after two learning trials, of the number of trials he would need to complete the learning task; (*b*) trials to mastery; (*c*) immediate recall—number of statements recalled immediately after learning; (*d*) delayed recall—number of statements recalled 24 hours after learning.

One major finding was that there was no overall difference on any of these four variables between those subjects in the high and low social desirability response style groups, but social desirability response style was related to all variables for women, but not men (i.e., women with high social desirability response style had higher confidence, learned in fewer trials, and had greater immediate and delayed recall).

Brown's statistical analysis makes it difficult to assess the meaning of this and other findings. It consisted of *t* tests of overall mean differences on each of the four learning measures between the high and low social desirability response-style groups within each sex group, and analyses of variance of overall differences among the four sex-social desirability response-style subject groups and among the four social desirability-descriptiveness item types, and the interaction between the subject and item type variables. Hence, it is not possible to assess directly the independent effects of each of the two subject variables or each of the two-item variables or the interactions among these four variables. Since Brown's findings are important and interesting, his data were reanalyzed to evaluate these issues.

---

[1] Since only an overall confidence rating, without respect to the characteristics of the statements of the EPPS, was obtained, assessment of the confidence associated with different item types is precluded.

## METHOD

The raw data reported by Brown (1957) for six variables were reanalyzed—the four learning measures analyzed by Brown (1960) as well as two other learning measures which were originally investigated but not analyzed—blockings (the number of trials between initial success and final mastery of an item) and time needed for mastery. Learning time and level of confidence were each analyzed in separate 2 × 2 factorial design analyses of variance based on two between subject variables—sex and social desirability response style.[2] The other four variables—number of trials to mastery, number of blockings, immediate recall, and delayed recall—were analyzed in an extended Lindquist (1953) Type III design analysis of variance, based on two between-subject variables—sex and social desirability response style—and two within-subject variables—item social desirability and item descriptiveness. Transformations were applied to four variables to reduce heterogeneity of variance.

## RESULTS

The mean squares and *F* ratios for the analyses of variance of learning time and level of confidence appear in Table 1. These statistics for the other variables appear in Table 2.[3] Results are reported for both the raw and transformed data, but only the results for the transformed data are discussed for the variables which required transformation.

The overall difference between high and low social desirability response-style groups was significant ($p < .05$) for two dependent variables—confidence and delayed recall; those with high social desirability response style had greater confidence and recall. There was no overall significant ($p > .05$) difference between men and women on any of the six dependent variables.

Interactions between sex and social desirability response style were significant ($p < .05$) for three dependent variables: confidence, learning time, and immediate recall. These three interactions were produced by greater differences between the two women's groups than between the two men's groups, and opposite trends for men and women on two variables—learning time and immediate recall. Compared with the women with low social desirability response style, the women with high social desirability response style had significantly ($p < .05$) greater confidence and immediate recall, as determined by Scheffé's (1953) test. Despite the overall significant difference in confidence between high and low social desirability response-style groups, the two men's groups did not differ significantly ($p >$

TABLE 1

ANALYSIS OF VARIANCE OF CONFIDENCE AND LEARNING TIME

| Source | *df* | Confidence (Raw) | | Confidence (Log X) | | Learning time (In minutes) | |
|---|---|---|---|---|---|---|---|
| | | *MS* | *F* | *MS* | *F* | *MS* | *F* |
| SD style | 1 | 1081.600 | 3.78 | .504 | 6.52* | 562.500 | 2.23 |
| Sex | 1 | .000 | — | .117 | 1.52 | 122.500 | — |
| SD style × Sex | 1 | 1040.400 | 3.63 | .381 | 4.93* | 1440.000 | 5.71* |
| Within (error) | 36 | 286.456 | — | .077 | — | 252.083 | — |

* $p \leq .05$.

---

[2] Only overall measures, without respect to the characteristics of the statements of the EPPS, are available for these two dependent variables.

[3] Some of the statistics for the analysis of variance of trials to mastery differ slightly from the statistics reported by Brown (1960), evidently because of a minor typographical error in the tabulations of raw data in Brown's (1957) Table 19.

## TABLE 2

ANALYSIS OF VARIANCE OF TRIALS TO MASTERY, BLOCKINGS, IMMEDIATE RECALL, AND DELAYED RECALL

| Source | *df* | Trials to mastery (Raw) | | Trials to mastery (Log X) | | Blockings (Raw) | | Blockings (Log X + 1) | | Immediate recall (Raw) | | Immediate recall (Arcsin) | | Delayed recall (Raw) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *MS* | *F* | *MS* | *F* | *MS* | *F* | *MS* | *F* | *MS* | *F* | *MS* | *F* | *MS* | *F* |
| SD style | 1 | 709.806 | 1.47 | .048 | 1.11 | 21.756 | 2.17 | .123 | 1.14 | 3.906 | 2.78 | 2138.906 | 3.75 | 10.000 | 5.00* |
| Sex | 1 | .156 | — | .000 | — | 1.056 | — | .000 | — | .506 | — | 113.906 | — | .225 | — |
| Item descriptiveness (ID) | 1 | 162.006 | 2.36 | .015 | 3.12 | .306 | — | .018 | — | .756 | — | 68.906 | — | 1.600 | 1.68 |
| Item social desirability (ISD) | 1 | 897.756 | 13.07** | .073 | 14.97** | 33.306 | 3.64 | .413 | 4.28* | 5.256 | 6.02* | 2600.156 | 6.74* | 1.225 | 1.29 |
| SD style × Sex | 1 | 1398.306 | 2.90 | .108 | 2.49 | 29.756 | 2.97 | .362 | 3.34 | 7.656 | 5.46* | 3657.656 | 6.42* | 3.600 | 1.80 |
| ID × SD style | 1 | 13.806 | — | .001 | — | .056 | — | .018 | — | .006 | — | 12.656 | — | .625 | — |
| ID × Sex | 1 | .156 | — | .000 | — | 61.256 | 6.70* | .157 | 1.63 | .056 | — | 1.406 | — | .400 | — |
| ID × ISD | 1 | 45.156 | — | .003 | — | 24.806 | 2.71 | .165 | 1.71 | .056 | — | 68.906 | — | .100 | — |
| ISD × SD style | 1 | 24.806 | — | .001 | — | 20.306 | 2.22 | .205 | 2.13 | .006 | — | 1.406 | — | .900 | — |
| ISD × Sex | 1 | 2.256 | — | .000 | — | 7.656 | — | .005 | — | .306 | — | 68.906 | — | .625 | — |
| ID × ISD × SD style | 1 | 4.556 | — | .001 | — | .756 | — | .000 | — | .006 | — | 35.156 | — | .625 | — |
| ID × ISD × Sex | 1 | 47.306 | — | .003 | — | 10.506 | 1.15 | .065 | — | .156 | — | 68.906 | — | .900 | — |
| ID × SD style × Sex | 1 | 51.756 | — | .003 | — | 13.806 | 1.51 | .147 | 1.52 | 3.306 | 3.78 | 1182.656 | 3.06 | 1.225 | 1.29 |
| ISD × SD style × Sex | 1 | 6.006 | — | .001 | — | 28.056 | 3.07 | .250 | 2.59 | .156 | — | 35.156 | — | .100 | — |
| ID × ISD × SD style × Sex | 1 | 1.406 | — | .000 | — | 8.556 | — | .053 | — | .306 | — | 237.656 | — | .025 | — |
| $\text{Error}_1$[a] | 36 | 482.120 | | .043 | | 10.017 | | .108 | | 1.403 | | 570.156 | | 1.999 | |
| $\text{Error}_2$[b] | 108 | 68.692 | | .005 | | 9.147 | | .097 | | .874 | | 385.990 | | .950 | |

[a] $\text{Error}_1$ is the appropriate error term for SD style, Sex, and SD style × Sex.

[b] $\text{Error}_2$ is the appropriate error term for the other sources of variance.

* $p \leq .05$.

** $p \leq .01$.

.05) on this variable, learning time, or immediate recall.

Item social desirability was significant ($p < .05$) for trials to mastery, blockings, and immediate recall; the socially desirable items were learned in fewer trials, involved fewer blockings, and had greater immediate recall. Item descriptiveness was not significant ($p > .05$) for any dependent variable.

## DISCUSSION

The finding that social desirability response style was not related to all the learning measures for either the male or female groups, much less both groups, sharply limits the interpretation of these results. The sex differences in the relationship of social desirability response style and learning may have been produced, as Brown suggested, by interactions of the subjects or experimental conditions with the experimenter; there was only one experimenter, a male, in this study. This interpretation is supported by the mounting evidence of such experimenter-linked interactions in studies of anxiety and learning (Axelrod, Cowen, & Heilizer, 1956; Heilizer, Axelrod, & Cowen, 1956).

The present finding that social desirability response style was related to delayed recall for men and women, and confidence and immediate recall for women, and the finding by Jackson and Pacine (1961) that the tendency to make socially desirable responses to California Psychological Inventory items and college grade-point average were related lend some support to Edwards' (1957) reinterpretation of studies of the relationship of anxiety and learning.[4] Viewing such studies as demonstrating the generality of a tendency to make a good impression, he argues that those who tend to make a good impression on personality scales (e.g., by giving socially desirable responses), and hence report little anxiety on the Manifest Anxiety (*MA*) scale (Taylor, 1953), also make a good impression in learning situations (e.g., by learning faster, etc.).

Nonetheless, Edwards' reinterpretation does not adequately explain the results of many previous studies of anxiety and learning. Although it asserts that subjects who have high social desirability response style, and hence report little anxiety on the *MA* scale, should generally have superior learning, the bulk of the findings of previous studies (Sarason, 1960; Taylor, 1956) indicates that low anxious subjects only have superior performance on complex learning tasks (such as the one investigated by Brown); their performance is inferior on simple learning tasks. While this interaction between the *MA* scale and task complexity cannot be predicted by the social desirability response-style interpretation, it can be explained by current conceptualizations of anxiety as a drive variable and affective state.

Since social desirability response style, when it is conceptualized as a tendency to make a good impression, is not as adept as anxiety in accounting for existing findings, even though these two personality variables are presumed to be highly related at both the artifactual and substantive levels (Edwards, 1957), it may be more meaningful to consider this response style, at least on personality scales, as simply a reflection of anxiety.

---

[4] It should also be noted that two studies of verbal conditioning (Crowne & Strickland, 1961; Marlowe, 1962) found that high scorers on the Marlowe-Crowne *SD* scale (Crowne & Marlowe, 1960) were more conditionable than low scorers. These findings do not seem relevant to an evaluation of Edwards' conceptualization because the Marlowe-Crowne *SD* scale more resembles a Lie scale than an *SD* scale, and these two kinds of scales measure distinctly different variables (e.g., Jackson & Messick, 1962).

## REFERENCES

AXELROD, H. D., COWEN, E. L., & HEILIZER, F. The correlates of manifest anxiety in stylus maze learning. *J. exp. Psychol.*, 1956, **51**, 131–138.

BROWN, P. K. A study of the relationship between social desirability as a psychological variable and learning performance in a complex verbal task. Unpublished doctoral dissertation, University of Washington, 1957.

BROWN, P. K. The social desirability variable and verbal learning performance. *J. educ. Psychol.*, 1960, **51**, 52–59.

CROWNE, D. P., & MARLOWE, D. A new scale of social desirability independent of psychopathology. *J. consult. Psychol.*, 1960, **24**, 349–354.

CROWNE, D. P., & STRICKLAND, BONNIE R. The conditioning of verbal behavior as a function of the need for social approval. *J. abnorm. soc. Psychol.*, 1961, **63**, 395–401.

EDWARDS, A. L. *Manual for the Edwards Personal Preference Schedule.* New York: Psychological Corporation, 1953.

EDWARDS, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

HEILIZER, F., AXELROD, H. D., & COWEN, E. L. The correlates of manifest anxiety in paired associate learning. *J. Pers.*, 1956, **24**, 463–474.

JACKSON, D. N., & MESSICK, S. Response styles on the MMPI: Comparison of clinical and normal samples. *J. abnorm. soc. Psychol.*, 1962, **65**, 285–299.

JACKSON, D. N., & PACINE, L. Response styles and academic achievement. *Educ. psychol. Measmt.*, 1961, **21**, 1015–1027.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

MARLOWE, D. Need for social approval and the operant conditioning of meaningful verbal behavior. *J. consult. Psychol.*, 1962, **26**, 79–83.

SARASON, I. G. Empirical findings and theoretical problems in the use of anxiety scales. *Psychol. Bull.*, 1960, **57**, 403–415.

SCHEFFÉ, H. A method for judging all contrasts in the analysis of variance. *Biometrika*, 1953, **40**, 87–104.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

TAYLOR, JANET A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303–320.

(Received July 19, 1962)

*Journal of Educational Psychology*
1963, Vol. 54, No. 1, 57-62

# PERCEPTUAL RECOGNITION OF APPROXIMATIONS TO ENGLISH IN RELATION TO SPELLING ACHIEVEMENT[1]

MICHAEL A. WALLACH
*Duke University*

In an experiment with 55 5th-grade children, ease of perceptual recognition was determined for nonsense words which resemble English and for nonsense words which do not. With all children being equated in recognition accuracy for the latter words, good spellers were found to recognize the former words much more readily than poor spellers. These and further results permitted the conclusion that good spellers show much greater transfer of training than poor spellers—whether the basis for this transfer rests on the learning of the sequential probability structure of letters, on phonetic generalization, or on both.

The present study bears on the question of whether spelling achievement may importantly depend on transfer of training. The basis for such transfer may rest on learning the general sequences in which letters tend to fall in the language, on phonetic generalization, or on both. To the extent that good spelling depends on such transfer of training, rather than just on learning the specific letter and sound sequences involved in particular words that are studied, nonsense words which resemble English, such as APHYST or EDITOL, should be more familiar for good spellers than for poor. Since familiarity is a powerful determinant of ease of perceptual recognition when a word is exposed for a brief period of time (Howes & Solomon, 1951), good spellers hence should recognize such nonsense words more easily, when they are presented tachistoscopically, than do poor spellers.

The present experiment tests predictions which hence are derived from the hypothesized importance for spelling of phonetic transfer—transfer of sound-letter rules based on the pronounceability of sound sequences—and of learning the sequential probability structure of letters. While these two factors may turn out to be related, a focusing of attention on the latter suggests some new implications regarding what one should attempt to teach in order to improve a child's spelling. Whereas teaching procedures based on phonetic training have been in use for some time, there has been little attempt yet to base training procedures on transfer involving the sequential probability structure of letters. These implications will be explored later in the paper.

## METHOD

### *Subjects*

The subjects were 55 fifth grade children from a public school in suburban Boston, approximately equally divided between the sexes.[2]

---

[1] The research reported herein was supported in part through Grant M-1324 from the National Institute of Mental Health, Public Health Service, and in part through the Cooperative Research Program of the Office of Education, United States Department of Health, Education, and Welfare. The former grant was under the general direction of J. S. Bruner, whose encouragement and aid are gratefully acknowledged. Thanks also are due W. D. Rohwer, Jr., for his assistance in conducting the study. Preliminary results which provided the impetus for the present investigation were obtained by W. Hull.

[2] Thanks are due E. R. Caverly, Superintendent of Schools, J. R. Hobson, Director of Child Placement, and Elgie Clucas, Principal

Since no sex differences were found in any of the correlational results in the study, the data for both sexes were pooled. All the children were of middle or upper-middle class socio-economic status.

### General Procedure and Predictions

The experimental procedure required first that we equate all children with regard to degree of perceptual recognition accuracy for nonsense words consisting of completely random letter sequences. This was achieved by setting the length of tachistoscopic exposure for each subject at the duration required for him to recognize approximately 33% of the letters in such words correctly. The basic perceptual measure then consisted in the degree of increase in recognition accuracy exhibited for nonsense words consisting of familiar letter sequences. Our major prediction was that this measure of degree of increase in recognition accuracy would be strongly correlated with spelling achievement. We expected further that degree of increase in recognition accuracy would be *less* strongly correlated with achievement indices in fields other than spelling, and with general intelligence. In other words, we predicted a relationship between recognition accuracy for nonsense words that resemble English and spelling achievement that would be specific to spelling achievement in at least some degree.

### Perceptual Recognition Session

*Stimulus materials.* Two main kinds of six-letter words were tachistoscopically exposed to each child: words of zero-order and fourth-order approximation to English. The words themselves were adapted from ones used by Miller, Bruner, and Postman (1954). That reference and Shannon (1948) describe in detail their method of construction.

In brief, nonsense words of zero-order approximation to English are constructed by a random sampling of the letters in the alphabet. Every letter in the alphabet hence has an equal chance of occurrence in each position. The result is words such as GFXRWM, VQWBVI, and CVGJCD.

Nonsense words of fourth-order approximation to English, on the other hand, are extremely close in nature to actual words in the language. While neither fourth-order words nor zero-order words can be found in the dictionary, fourth-order words look as if they could be English while zero-order words do not. Words of fourth-order approximation to English are constructed by taking account of the sequences in which strings of four letters occur in the language. Beginning at random in English prose with a three-letter sequence (such as TER), we search the continuation of the prose until this three-letter sequence occurs again, and then add the letter that follows it (yielding, for example, TERV). Continuing now with the last three letters (ERV), we proceed to read the prose until we find this new three-letter sequence again, and we once again add the letter that follows it (the result may, for instance, be ERVA). Proceeding again with the last three letters (RVA), we search for its reoccurrence and add the letter that follows it (for instance, RVAL). The six-letter word is now complete, and reads TERVAL. To construct another such word, we again begin at random in English prose with a new three-letter sequence. The result of this procedure is words such as MOSSIA, PRESPE, and APHYST.

Thus, the fourth-order words are highly similar to English, while the zero-order words are completely random strings of letters.

A third kind of nonsense words, first-order approximations, were also included, to serve as a control. Nonsense words of first-order approximation to English are constructed by a random sampling of letters from English prose. These words hence contain letters in relative frequencies proportional to their likelihood of occurrence in the language. The *sequential* structure of these letters, however, still is essentially random. Examples are DNEHHS and WDNMII. It was our expectation that sensitivity to mere relative frequency of occurrence per se of letters in English prose would not be relevant to good spelling. Rather, we predicted that it was sensitivity to the transitional probabilities governing the arrangement of these letters into sequences that specifically underlay good spelling. The first-order words hence were included as a control and were expected to yield results very close to those found for zero-order words.

Table 1 lists all stimulus words used in the study, in their order of presentation. The eight zero-order words listed under Part I of the table constituted the training series. The sequence shown was presented two or three times, with the experimenter varying the tachistoscopic shutter speed until the recognition accuracy criterion of 33% of the letters correct

---

of the Michael Driscoll School in Brookline, Massachusetts, for permission to work with the pupils of the Michael Driscoll School in carrying out the present research.

TABLE 1

STIMULUS WORDS IN THEIR SEQUENCE OF PRESENTATION

Part I: Training series

| | Word | Order of approximation |
|---|---|---|
| 1. | PMZCEN | 0 |
| 2. | OYOEVZ | 0 |
| 3. | IJHBWS | 0 |
| 4. | BHDTUN | 0 |
| 5. | MFRSIW | 0 |
| 6. | HHJHUF | 0 |
| 7. | GFXRWM | 0 |
| 8. | YRULPZ | 0 |

Part II: Test series

| Word | Order | Word | Order | Word | Order |
|---|---|---|---|---|---|
| 1. MOSSIA | 4 | 9. GFUJXZ | 0 | 17. WDNMII | 1 |
| 2. CVGJCD | 0 | 10. FOREMS | 4 | 18. TATION | 4 |
| 3. VTYEHU | 1 | 11. OZHGPM | 0 | 19. WXPAUJ | 0 |
| 4. ONETIC | 4 | 12. DNEHHS | 1 | 20. EDITOL | 4 |
| 5. EVERAL | 4 | 13. IDCFRA | 1 | 21. TERVAL | 4 |
| 6. YDEWAK | 1 | 14. LYMIST | 4 | 22. AAESPW | 1 |
| 7. VORIAL | 4 | 15. KERSON | 4 | 23. PRESPE | 4 |
| 8. APHYST | 4 | 16. EGQMNW | 0 | 24. VQWBVI | 0 |

was achieved over a block of four successive words. The length of exposure that met this criterion varied from 1/200 second to 1/5 second for different subjects.

The 24 words listed under Part II of the table constituted the test series. The sequence shown was presented two times, with all 48 presentations for a given subject being shown at the exposure speed which had met the 33% recognition accuracy criterion in Part I. Of the 24 words in Part II, 12 were fourth-order approximations to English, 6 were zero-order, and 6 were first-order. Each block of 12 words included half the words of each order.

*Apparatus, preliminary familiarization, and response materials.* A Bell and Howell slide projector was used in conjunction with an Ace tachistoscopic shutter. The image was projected on a milk glass screen which the child viewed from the other side. The screen was mounted in a wooden cabinet. The six-letter words cast an image 2 inches long and ½ inch high on the screen. The eyes of the viewer were about 3 feet away. Uniform fixation was insured by boundary lines which outlined the rectangular space within which each word was flashed.

Prior to the training series and test series, each child was familiarized with the general nature of the tachistoscope by flashing English words at varying exposure lengths. The words used in this preliminary familiarization were LOVELY, MINUTE, BOTTOM, and WINTER.

Responses to the stimuli in the training and test series were written by the child in a booklet of blank sheets that he kept in front of him. Each sheet of the booklet had room for eight words to be filled in. The spaces where the letters of each word were to be entered were indicated by a row of six dashes. The child was instructed always to fill in all six spaces for each word that was flashed, guessing whenever necessary.

*Perceptual measure.* Degree of increase in recognition accuracy for potentially familiar in contrast to unfamiliar nonsense words was determined through a comparison of results for the fourth-order and the zero-order words in the test series.

The number of letters correct and correctly placed was calculated for the presentations of zero-order words in the test series. Since there was a total of 12 presentations of zero-order words (6 words shown twice each), there hence was an opportunity for 72 correct letters. In the case of all 55 subjects, there was success in achieving the criterion of 33% of the zero-order letters correct in the test series, plus or minus 8% as an allowance for accidental variation from the criterion.[3] All 55 subjects thus obtained from 18 to 30 of the 72 letters correct, and hence were approximately matched in terms of their recognition accuracy level for zero-order words.

The number of letters correct and correctly placed was likewise calculated for the fourth-order words. Since there was a total of 24 presentations of fourth-order words (12 words shown twice each), there was an opportunity for 144 correct letters. Given the fact of twice as many fourth-order as zero-order words, the number of letters correct for the zero-order words was doubled in the case of each child.

A ratio score was computed for each subject, consisting of the number of fourth-order letters correct minus twice the number of zero-order letters correct, divided by twice the number of zero-order letters correct. This ratio score may be expressed:

$$\frac{\Sigma 4\text{th order} - 2(\Sigma 0\text{ order})}{2(\Sigma 0\text{ order})}$$

This measure, in a percentage, indicated the degree of increase in perceptual recognition accuracy for the fourth-order words, all subjects being equated at approximately 33% recognition accuracy for zero-order words as the baseline of comparison.

---

[3] A few individuals who fell outside this 8% error allowance were discarded from the study prior to analysis of the results. A subsequent check indicated that their inclusion would not have altered any of the findings.

Comparisons between fourth-order and first-order words, and between first-order and zero-order words, also were carried out for control purposes.

### *Spelling Achievement, Achievement in Other Areas, and Intelligence*

Spelling achievement was assessed from the spelling achievement score in the Intermediate Battery of the Metropolitan Achievement Tests (Allen, Bixler, Conner, & Graham, 1946). Achievement in three other areas also was assessed: reading, English, and arithmetic. The average reading achievement score, English achievement score, and average arithmetic achievement score, all again from the Intermediate Battery of the Metropolitan Achievement Tests, provided the measures. General intelligence, finally, was assessed from the Kuhlmann-Anderson (1942) IQ. The IQ was chosen as the general intelligence measure in order to provide an assessor that would be as standard, general, and familiar in its psychological significance as possible. For, in the case of intelligence, our specific interest was in an overall evaluation of ability. The Metropolitan Achievement Tests (Intermediate Battery) and the Kuhlmann-Anderson Intelligence Tests were administered to all of our fifth grade subjects during the spring term of the fifth grade year.

## RESULTS AND DISCUSSION

The ratio scores indicating degree of increase in perceptual recognition accuracy for fourth-order as compared with zero-order words varied from 11.9% improvement to 197.2% improvement, i.e., from almost no improvement at all to almost 200% improvement. The first and major prediction of the experiment concerned the relationship between these scores and spelling achievement scores. As indicated in Table 2, the product-moment correlation coefficient between the two variables is .47, which is significant well beyond the .001 level of confidence for the sample size of 55. Our next prediction concerned the degree of specificity of this relationship. We find from Table 2 that the perceptual measure is not as strongly related to average reading achievement scores, although this relationship is still significant ($r=.37$, which is significant beyond the .01 level). We also note that the perceptual measure is still less strongly related to English achievement, and even less strongly related to average arithmetic achievement ($r=.33$ and .30, respectively, both of which are significant beyond the .05 level). Finally, we find that the perceptual measure is not significantly related to general intelligence ($r=.23$).

TABLE 2

PRODUCT-MOMENT CORRELATION COEFFICIENTS BETWEEN DEGREE OF INCREASE IN RECOGNITION ACCURACY FOR FOURTH-ORDER AS COMPARED WITH ZERO-ORDER WORDS, AND VARIOUS MEASURES OF INTELLECTUAL PERFORMANCE

| Measure | $r$ |
|---|---|
| Spelling achievement [a] | .47*** |
| Average reading achievement [a] | .37** |
| English achievement [a] | .33* |
| Average arithmetic achievement [a] | .30* |
| IQ [b] | .23 |

Note.—$N = 55$.

[a] Metropolitan Achievement Tests: Intermediate Battery (Allen et al., 1946).

[b] Kuhlmann-Anderson Intelligence Tests (Kuhlmann & Anderson, 1942).

* $p < .05$.

** $p < .01$.

*** $p < .001$.

The question of whether the various correlations we have reviewed are reliably different from each other was investigated by means of a technique developed by Hotelling (see Walker & Lev, 1953, pp. 256–257). By a two-tailed $t$ test (which actually is conservative since our prediction specifies direction), the correlation of .47 between the perceptual measure and spelling achievement is different from the correlation of .23 between the perceptual measure and general intelligence at almost the .05 level ($t=2.00$). The .47 correlation is not significantly different from the other correlations by Hotelling's technique, however.

Partial correlation coefficients, in turn, were computed between the perceptual measure and spelling achievement, holding statistically constant the reading, English, arithmetic, and IQ scores in succession. In every case the partial correlation coefficient remained significant at the .05 level or better.

Not only, therefore, is there a strong relationship between the perceptual measure and spelling achievement, but also this relationship is specific to spelling achievement in at least some degree. The perceptual measure is less strongly related to indices of verbal achievement in areas other than spelling, is still less strongly related to achievement in the numerical sphere, and is not significantly related to a general omnibus intelligence score. Furthermore, even when each of these latter indices is held constant statistically, there remains a significant relationship between the perceptual measure and spelling achievement.

Regarding the perceptual findings for first-order words, the following results were obtained. Ratio scores based on degree of increase from first-order to fourth-order words yielded correlation results essentially identical to those just described for ratio scores based on degree of increase from zero-order to fourth-order words. In turn, ratio scores based on degree of increase from zero-order to first-order words yielded nonsignificant correlations in all cases. These control results demonstrate that sensitivity to mere relative frequency of occurrence per se of letters in English does not underlie spelling achievement. Rather, it is sensitivity to the transitional probabilities that characterize the sequential ordering of these letters that specifically underlies spelling achievement.

We may conclude that the degree of increase in perceptual recognition accuracy for nonsense words whose letters constitute close approximations to the sequential structure of English, as compared with nonsense words whose letters are arranged in random sequences, is strongly correlated with spelling achievement. In addition, this correlation possesses a goodly degree of specific connection with the spelling achievement variable. These results are consistent with the proposition that good spellers, in contrast to poor spellers, are learning not just to spell particular words, but rather also are learning a much more general property of English words—namely, the transitional probability structure governing the sequential arrangement of letters in such words. The present results also are consistent with the proposition that spelling achievement is dependent upon phonetic transfer or phonetic generalization — transfer of sound-letter rules based on the ease of pronunciation of sound sequences.

Whether or not the two proposals turn out to be related, however, consideration of the findings of the experiment from the viewpoint of the sequential probability structure of letters suggests some new implications that may be of relevance for the teaching of spelling. It is consistent with the results of the present experiment to suggest that one problem for poor spellers—or for some poor spellers—may be their lack of internalization of the transitional probability structures that characterize letter sequences in English. One pedagogical approach to this problem might be to present poor spellers with nonsense words of varying degrees of approximation to English and reward them for correctly discriminating the ones that are most like English. This would be a "concept attainment" type of procedure in which the correct instances of the concept in question are nonsense words of fourth-order approximation to English. As poor spellers become more proficient at attaining this concept, they will be learning to recognize the general letter sequences that characterize English words. This new knowledge, in turn, then may aid their spelling of particular English words.

The present experiment strongly supports the conclusion that good spellers show much greater transfer of training than poor spellers—whether the basis for this transfer depends on the learning of the sequential probability structure of letters, on phonetic generalization, or on both.

## REFERENCES

ALLEN, R. D., BIXLER, H. H., CONNER, W. L., & GRAHAM, F. B. *Metropolitan Achievement Tests: Intermediate Battery.* Yonkers-on-Hudson, N. Y.: World Book, 1946.

HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word-probability. *J. exp. Psychol.,* 1951, 41, 401–410.

KUHLMANN, F., & ANDERSON, ROSE G. *Kuhlmann-Anderson Intelligence Tests.* Minneapolis: Educational Pubishers, 1942.

MILLER, G. A., BRUNER, J. S., & POSTMAN, L. Familiarity of letter sequences and tachistoscopic identification. *J. gen. Psychol.,* 1954, **50,** 129–139.

SHANNON, C. E. A mathematical theory of communication. *Bell Sys. tech. J.,* 1948, **27,** 379–423, 623–656.

WALKER, HELEN M., & LEV, J. *Statistical inference.* New York: Holt, 1953.

(Received July 30, 1962)

*Journal of Educational Psychology*
1963, Vol. 54, No. 1, 63–71

# DIFFERENTIAL COLLEGE EFFECTS ON THE MOTIVATION OF TALENTED STUDENTS TO OBTAIN THE PHD[1]

ALEXANDER W. ASTIN
*National Merit Scholarship Corporation, Evanston, Illinois*

An attempt was made to identify differential college effects on the student's motivation to seek the PhD degree. An input-output design for controlling differential student input was applied in a 4-year longitudinal study of 6544 high aptitude students. PhD aspiration appeared to be negatively affected by the size of the student body, the percentage of males in the student body, and the Conventional Orientation in the college environment. The Social Orientation also affected PhD aspiration significantly, but the relationship was very complex. Attending a coeducational liberal arts college appears to increase the student's motivation to seek the PhD, whereas PhD aspiration tends to be reduced as a function of attending one of the northeastern men's colleges.

Studies of differential college effects are difficult to design, primarily because students are not distributed randomly among institutions (Astin & Holland, 1962; Heist, 1960; Holland, 1959). Consequently, variations in student performance from one college to another probably reflect, at least to some extent, differences in their student inputs. This problem is further complicated by the fact that the independent variable—the college environment—is also related to the student input (Astin & Holland, 1961). Under such conditions, the investigator runs the risk of finding "effects" of college environments which are in fact due to variations in student inputs.

One possible approach to controlling student input is partial correlation. Correlations between the college environment and the student performance are obtained, and the contribution of relevant input variables partialed out (e.g., Astin, 1961). However, if the magnitude of the environmental effect is proportional to the institution's score on the input variable, then the effect will account for some or all of the correlation between input and student performance. Under such conditions, and when the institution is used as the unit of sampling, partial correlation is capable of partialing out the institutional effect along with the input variable.

Another approach to the problem of differential student input is to match students on the relevant input variables. But matching is subject to another kind of difficulty, namely, the regression effect which occurs when subsamples of students are selected which are unrepresentative of their respective college populations. This regression effect could tend either to produce spuriously positive environmental "influences," or to mask real environmental effects.

The present study attempted to compare the effects of different college characteristics on the student's motivation to obtain the PhD degree. However, as an alternative approach to partial correlation or matching, this project employed a modification of an input-output design used in a recent study of "PhD productivity" (Astin, 1962c). This design begins by pooling all subjects without respect to college attended. On the basis of the relevant input variables, an actuarial table is generated in which the cell entry represents an "expected" score on the student performance. These expected scores are obtained from a new sample of subjects, and the discrepancy between expected and actual performances calculated. This discrepancy score is then used as the dependent variable in the analysis of college effects.

---

[1] This study is a part of the research program of the National Merit Scholarship Corporation, and was supported by a grant from the National Science Foundation. The author is indebted to R. C. Nichols for his many valuable suggestions.

This method, which will be outlined in more detail in the following sections, is analogous to the analysis of covariance. It has the advantage of being able to deal with nonlinear relationships, categorical predictors, and interactions among predictors, but it is more limited in that it requires large numbers of subjects.

## METHOD

### *Subjects*

The sample of 6,544 Merit Finalists and recipients of the Letter of Commendation from the 1957 National Merit Scholarship competition included 4,374 males and 2,170 females. These subjects had completed a mailed questionnaire in the Fall of 1957, shortly after enrolling for their freshman year in college. The input (predictor) data were obtained at that time. A follow-up questionnaire was administered to these students in the Spring of 1961, approximately at the expected time of their graduation from college. Since 80% of the students returned questionnaires in 1957, and 78% of them also returned questionnaires in 1961, the 6,544 subjects comprise about 62% of the original sample.

### *Criterion Measure*

In the follow-up questionnaire the students were asked to indicate the highest degree which they were seeking (MA, PhD, etc.). Since most previous studies of college effects had used attainment of the PhD as the dependent variable, the student's highest degree sought after 4 years of college was dichotomized (PhD) versus non-PhD), to make the results more comparable. All other degrees, including the professional degrees, were considered as "non-PhDs." It was also felt that dichotomizing the student's level of aspiration (LOA) in this manner would avoid some of the logical problems connected with scaling or rank ordering the various degrees (Astin, 1962b).

### *Procedure*

*Cross-validation.* From the total sample of 6,544 students, a cross-validation sample of 1,908 students was selected for the analyses of college effects. The criterion in selecting these students was that each attended one of 82 institutions which enrolled 15 or more students from the total sample in the Fall of 1957. In order to minimize the contribution of any single college to the effects analyses, not more than 25 students from each college were chosen. If a college enrolled more than 25 of the 6,544 students in 1957, the 25 students for the cross-validation sample were selected at random.

The 82 institutions attended by these cross-validation subjects were rank-ordered on the following college variables: Affluence, Size, Masculinity, Private (versus Public) Control, Homogeneity, Realistic Orientation, Intellectual Orientation, Social Orientation, Conventional Orientation, Enterprising Orientation and Artistic Orientation. These college variables include the eight scales from the Environmental Assessment Technique (EAT; Astin & Holland, 1961), in addition to scores on three other factors recently identified in an analysis of 33 college variables (Astin, 1962a). On the basis of their rank order, institutions were grouped into seven levels of about equal size separately on each college variable. (The Private-Public variable, of course, had only two levels.) Through this procedure approximately 275 students from 11 or 12 different institutions were assigned to each of the seven levels of a college variable.

*Weighting the input variables.* The remaining 4,636 students were used to determine weights for the following input variables:

1. LOA (highest degree sought) in 1957
2. Sex
3. Major field of study in 1957
4. Career choice in 1957
5. Degree of satisfaction with career choice in 1957 (dichotomized as "very satisfied" versus all other degrees of satisfaction)
6. Verbal Aptitude (Scholarship Qualifying Test)
7. Mathematical Aptitude (Scholarship Qualifying Test)

8. High school percentile rank
9. High school curriculum (academic, commercial, etc.)
10. Number of high school math courses taken
11. Number of scholarships applied for in 1957
12. Number of scholarship offers received in 1957
13. Distance from home to college
14. Commuter versus resident in 1957
15. Father's educational level
16. Mother's educational level
17. Father's attitude toward student's attending college (four-point scale: "Insists that I go," "Wants me to go," "Indifferent," "Doesn't want me to go")
18. Mother's attitude toward student's attending college (same as Number 17)
19. Number of books in the home
20. Percentage of high school peers attending college
21. Father's occupation

The relationships between each of these 1957 input variables and the probability of aspiring to the PhD in 1961 were determined separately by means of actuarial tables. Beginning with the most efficient predictor, other input variables were added to each cell of the actuarial table until either (*a*) additional variables no longer produced significant changes in PhD probabilities; or (*b*) the $n$ in the particular cell became too small for reliable discriminations. For example, it was found that sex and initial LOA (degree sought in 1957) were the two most effective predictors of 1961 PhD aspiration. Accordingly, the subject were first separated into cells on the basis of their sex and 1957 LOA. Then, separately within each sex × 1957 LOA cell, all other input variables were applied to determine if further discriminations in terms of 1961 PhD aspiration were possible. The subjects in cells with relatively large $n$'s could usually be classified on a greater number of additional input variables than the subjects in cells with relatively small $n$'s.[2] The total number of input variables finally used to classify each subject ranged from two to five.

### *Analyses of College Effects*

Students in the cross-validation sample were classified on the same input variables used in the final actuarial table. All cross-validation subjects in a particular cell were assigned an input "score" equal to the proportion of PhD aspirants in that cell among the subjects from the original sample. This value was designated as the subject's probability of aspiring to the PhD after 4 years of college. To obtain discrepancies between these expected values and the subject's actual performance, a score of 1.00 was given to those cross-validation subjects who actually aspired to the PhD in 1961, and a score of 0 to those who did not aspire to the PhD in 1961. Each subject's expected score (probability from the original actuarial table) was subtracted from his criterion score (1.00 or 0), yielding a discrepancy score. This discrepancy score was used as the dependent variable in the analyses of college effects.

In assessing the influence of a given college variable on the student's motivation to obtain the PhD, the object was to determine if the mean discrepancy score differed as a function of different levels on the college variable. However, since the distribution of discrepancy scores was not normal, an analysis of variance of discrepancy scores between levels of the college variable was not considered to be an appropriate statistical technique. As an alternative procedure, the Kruskal-Wallis $H$ statistic was employed. The $H$ is approximately 95% as efficient as the single classification analysis of variance (Seigel, 1956), but does not require normality or homogeneity of variance in the dependent variable.

## RESULTS

Table 1 summarizes the final actuarial table. Each row in Table 1 represents a cell in the table (i.e., a subgroup of subjects classified in given ways on certain input variables). The most potent predictors of later PhD aspiration turned out to be the student's initial LOA, sex, and career choice. However, even when these three variables are controlled, the student's degree of satisfaction with his vocational choice, his major field, his verbal aptitude, whether or not he holds a scholarship, and his father's level of education are still related to later PhD aspiration in some of the subgroups.

---

[2] The .10 level of confidence was generally used in selecting additional input variables. However, to reduce the likelihood of Type II errors (which in controlling student input would be the more serious kind), this level was exceeded in a few instances where the cell $n$ was extremely small.

TABLE 1

PROBABILITY OF 1961 PHD ASPIRATION FOR SUBGROUPS CLASSIFIED ON 1957 INPUT VARIABLES

| 1957 input variable | | | | | | | | Actuarial table (*N* = 4,636) | | Cross-validation dnoɹ⅁ (*N* = 1,908) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Initial LOA | Sex | Career choice [a] | Satisfied with choice? | Major field | Scholarship held? | Verbal aptitude | Father's educational level | *n* | *p*** | n | p** |
| < Bachelor's | F | — | — | — | — | — | — | 49 | .06 | 8 | .25 |
| Bachelor's | F | SR or CT | — | Chemistry | — | — | — | 43 | .26 | 20 | .15 |
| Bachelor's | F | SR or CT | — | All others | — | — | — | 37 | .10 | 17 | .12 |
| Bachelor's | F | All others | — | — | Yes | High | — | 131 | .16 | 42 | .14 |
| Bachelor's | F | All others | — | — | Yes | Middle | — | 219 | .06 | 70 | .10 |
| Bachelor's | F | All others | — | — | Yes | Low | — | 35 | .06 | 5 | .00 |
| Bachelor's | F | All others | — | — | No | High | — | 86 | .13 | 30 | .17 |
| Bachelor's | F | All others | — | — | No | Middle | — | 175 | .05 | 68 | .09 |
| Bachelor's | F | All others | — | — | No | Low | — | 34 | .00 | 13 | .08 |
| Bachelor's | M | SR or CT | — | Chemistry | — | — | — | 15 | .67 | 8 | .50 |
| Bachelor's | M | SR or CT | — | All others | — | — | — | 25 | .20 | 8 | .50 |
| Bachelor's | M | E | — | — | No | — | — | 97 | .22 | 41 | .22 |
| Bachelor's | M | All others | — | — | No | — | — | 133 | .11 | 71 | .18 |
| Bachelor's | M | All others | — | — | Yes | High | — | 83 | .35 | 40 | .28 |
| Bachelor's | M | All others | — | — | Yes | Middle | — | 194 | .23 | 87 | .18 |
| Bachelor's | M | All others | — | — | Yes | Low | — | 39 | .13 | 21 | .05 |
| Master's | F | SR or CT | — | — | Yes | — | — | 90 | .30 | 37 | .41 |
| Master's | F | SR or CT | — | — | No | — | — | 42 | .17 | 30 | .27 |
| Master's | F | U | — | — | — | — | — | 69 | .41 | 32 | .59 |
| Master's | F | All others | — | — | — | — | — | 295 | .16 | 113 | .13 |
| Master's | M | SR or CT | Yes | — | — | — | — | 77 | .49 | 26 | .46 |
| Master's | M | SR or CT | No | — | Yes | — | — | 28 | .43 | 10 | .40 |
| Master's | M | SR or CT | No | — | No | — | — | 58 | .66 | 24 | .54 |
| Master's | M | E | — | — | Yes | — | — | 248 | .34 | 103 | .29 |
| Master's | M | E | — | — | No | — | — | 151 | .21 | 52 | .37 |
| Master's | M | HST or SW | — | — | — | — | — | 47 | .49 | 8 | .50 |
| Master's | M | U | — | — | — | — | — | 108 | .41 | 45 | .47 |
| Master's | M | All others | — | — | — | — | — | 169 | .30 | 57 | .26 |
| PhD | F | SR or CT | — | — | — | — | — | 111 | .63 | 54 | .65 |
| PhD | F | All others | — | — | — | — | — | 56 | .50 | 23 | .52 |
| PhD | M | SR or CT | Yes | — | — | — | College + | 176 | .77 | 88 | .76 |
| PhD | M | SR or CT | Yes | — | — | — | < College | 217 | .68 | 89 | .74 |
| PhD | M | SR or CT | No | — | — | — | College + | 120 | .69 | 75 | .80 |
| PhD | M | SR or CT | No | — | — | — | < College | 147 | .51 | 54 | .68 |
| PhD | M | E | — | — | — | — | College + | 61 | .59 | 41 | .51 |
| PhD | M | E | — | — | — | — | < College | 73 | .49 | 44 | .34 |
| PhD | M | U | — | — | — | — | — | 150 | .51 | 68 | .65 |
| PhD | M | All others | — | — | — | — | — | 109 | .63 | 31 | .42 |
| Professional | F | — | — | — | — | — | — | 71 | .30 | 28 | .14 |
| Professional | M | — | — | Science | Yes | — | — | 155 | .19 | 64 | .14 |
| Professional | M | — | — | Nonscience | Yes | — | — | 138 | .34 | 59 | .33 |
| Professional | M | — | — | — | No | — | — | 201 | .16 | 80 | .23 |
| Unclassified | F | — | — | — | — | — | — | 31 | .10 | 6 | .17 |
| Unclassified | M | — | — | — | Yes | — | — | 20 | .50 | 11 | .45 |
| Unclassified | M | — | — | — | No | — | — | 23 | .26 | 7 | .29 |

[a] Career choices have been abbreviated as follows: SR, Scientific Researcher; CT, College Teacher; E, Engineer; HST, High School Teacher; SW, Social Worker; U, Undecided.

[b] Proportion of subjects aspiring to the PhD in 1961.

The cell with the highest rate of PhD aspiration in 1961 (77%) consisted of 176 males who initially aspired to the PhD in 1957, who had chosen either college teaching or scientific research for a career, who were well satisfied with their career choice, and whose fathers were themselves college graduates. The cell with the lowest rate of PhD aspiration in 1961 (0%) consisted of 34 females who initially aspired to the bachelor's degree, who had chosen a career other than either scientific research or college teaching, who were of relatively low verbal aptitude, and who did not hold a scholarship. (These two extreme groups had 1961 PhD aspiration rates of 76% and 08%, respectively, in the independent cross-validation sample.)

Figure 1 shows the input probabilities from the initial sample applied to the entire cross-validation sample. Distributions of probabilities clearly separate those who actually aspired to the

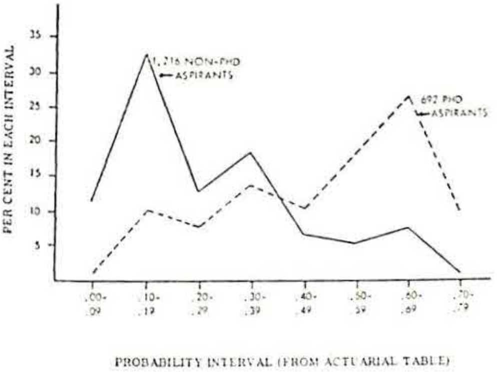


FIG. 1. Discrimination between PhD aspirants and non-PhD aspirants in the cross-validation sample.

PhD in 1961 from those who did not. If the a priori optimal cutting score of .50 is used, the overall accuracy of identification after 4 years is 74% in the cross-validation sample. Fifty-four percent of the PhD aspirants are correctly identified, and 84% of the non-PhD aspirants. Lowering the cutting score in the cross-validation sample to the a priori base rate (.33) correctly identifies 74% of the 1961 PhD aspirants and 65% of the 1961 non-PhD aspirants. The overall rate of accuracy, however, is reduced from 74% to 68% with this lowered cutting score.

Despite the relatively high relationships found between the input variables and later PhD aspiration, it is possible that certain other relevant input variables may not have been controlled. Since the 21 variables which were tried are primarily concerned with educational and vocational plans, aptitude, and socioeconomic factors, perhaps the most likely sources of uncontrolled input variance are personality variables. A chance to examine this possibility was afforded by the fact that, prior to entering college in 1957, about one-sixth of the subjects had taken two personality inventories: the California Psychological Inventory (CPI; Gough, 1957); and the Inventory of Beliefs (IB; Stern, Stein, & Bloom, 1956). Since Holland (1959) had shown previously that several scales from the CPI and IB are related to characteristics of the college chosen by the student, it was considered particularly important to determine if these scales were also related to the dependent variable used in the present study.

Using the 304 cross-validation subjects for whom CPI and IB scores were available, the PhD discrepancy scores (1.00 or 0 minus the input probability) were normalized and correlated with each of the inventory scales. None of the 20 coefficients (18 for the CPI and 2 for the IB) was significant ($p = .05$), results which suggest that personality, in CPI and IB terms, was not a major source of uncontrolled input variance.

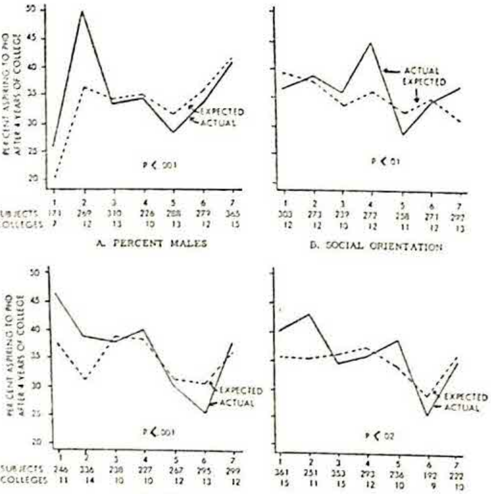
Results of the analyses of effects of the 12 college variables are presented in Table 2. Four of these variables (Size, Percentage of Males, the Social Orientation, and the Conventional Orientation) appear to have significant effects on the student's PhD aspiration. Plots of these

TABLE 2

ANALYSES OF COLLEGE EFFECTS ON PHD ASPIRATION

| College variable | *df* | $\chi^2$ (Kruskal-Wallis *H*) |
|---|---|---|
| Affluence | 6 | 4.64 |
| Size (Total Enrollment) | 6 | 26.31*** |
| Maculinity I[a] | 6 | 7.78 |
| Masculinity II (Percentage of Males) | 6 | 37.29*** |
| Homogeneity | 6 | 12.58 |
| Private versus Public | 1 | 2.12 |
| Realistic Orientation | 6 | 11.31 |
| Intellectual Orientation | 6 | 5.68 |
| Social Orientation | 6 | 21.00** |
| Conventional Orientation | 6 | 15.77* |
| Enterprising Orientation | 6 | 4.88 |
| Artistic Orientation | 6 | 9.74 |

[a] The Conventional Orientation minus the Social and Artistic Orientations (from Astin, 1962a).
* $p \leq .02$.
** $p \leq .01$.
*** $p < .001$.



A. PERCENT MALES

B. SOCIAL ORIENTATION

C. SIZE

D. CONVENTIONAL ORIENTATION

FIG. 2. Expected and actual percentages of PhD aspirants at each of seven levels on four college variables.

four variables are shown in Figure 2. Expected percentages of PhD aspirants for each of the seven levels of the four college variables have been computed by averaging the expected probabilities for the subjects in each level and multiplying by 100. The magnitude of the effect at any level of a college variable is indicated by the *difference* between the expected percentage (broken line) and the actual percentage (solid line).

Size, Percentage of Males, and the Conventional Orientation tend to have a negative effect on PhD aspiration (i.e., expected percentages of PhD aspirants tend to exceed the actual percentages at higher levels on these variables, whereas at the lower levels the actual percentages exceed the expected percentages). However, the greatest differences between expected and actual percentages do not always occur in the extreme intervals of these college variables. Results with the variable, Percentage of Males in the Student Body, are of special interest. Colleges in both the first and second categories on this variable appear to have a facilitative effect on the student's PhD aspiration. However, the effect seems to be at least as pronounced in the second category (coeducational institutions with about equal proportions of males and females) as in the first category (women's colleges). The effect in this second category was similar for both sexes, with the actual PhD aspiration rates for both males and females being about 1.4 times their expected rates. The actual PhD aspiration rate for females in Category 1 (women's colleges) was about 1.3 times the expected rate. (It is important to note that adjusted PhD aspiration was not also related to Masculinity defined as the Conventional Orientation minus the Social and Artistic Orientations. Apparently the two different methods of scoring Masculinity are not equivalent.)

The Social Orientation appears to have a complex, nonmonotonic relationship to PhD aspiration, with the largest positive and largest negative effects occurring in adjacent intervals.

In view of the evidence that there are differential college effects on the student's PhD aspiration, an attempt was made to relate these findings to a recent study of PhD "productivity" (Astin, 1962c). In this latter study, the proportion of graduates at each of 265 institutions who later obtained the PhD was determined. These "actual" PhD outputs were then compared with "expected" PhD outputs based on three student input variables: sex, intelligence, and major field of study. Under these conditions, public institutions were found to be significantly "overproductive," and colleges for men located in the Northeastern states were significantly "underproductive."

From Table 2 it appears that the private and public institutions in the present study did not differ in the extent to which they affected the student's PhD aspiration. Possible reasons for this discrepancy in the results of the two studies are: (*a*) that a larger and more heterogeneous sample of institutions was used in the earlier study, (*b*) that differences in student input may not have been adequately controlled in the earlier study, or (*c*) that the student's stated PhD aspiration at college graduation does not predict attainment of the PhD with complete accuracy.

To re-examine the underproductivity of the Northeastern men's colleges reported in the earlier study, students attending these same colleges were identified in the cross-validation sample and compared with all other students by means of the Kruskal-Wallis analysis. Students attending two other subgroups of institutions—technological institutions and liberal arts colleges—were also identified and similar analyses performed. (Those latter two groups tended to be overproductive in the earlier study, although the differences were not statistically significant.) Table 3 shows the results of these additional analyses. The actual proportion of PhD aspirants at the Northeastern men's colleges fell short of the expected proportion ($p < .001$), a result which supports the finding of the earlier study. Technological institutions did not differ significantly from the others in the analysis, but the coeducational liberal arts colleges, when

TABLE 3

EFFECTS OF THREE DIFFERENT TYPES OF INSTITUTIONS ON THE STUDENT'S PHD ASPIRATION

| Type | Institution | Number of students | Expected percentage of PhD aspirants | Actual percentage of PhD aspirants | *df* | $\chi^2$ (Kruskal-Wallis *H*) |
|---|---|---|---|---|---|---|
| Northeastern men's colleges | Amherst<br>Dartmouth<br>Harvard<br>Princeton<br>Wesleyan<br>Williams<br>Yale | 167 | 41.28 | 34.13 | 1 | 14.22*** |
| Technological institutions | Auburn<br>California Institute of Technology<br>Case Institute of Technology<br>Georgia Institute of Technology<br>Illinois Institute of Technology<br>Massachusetts Institute of Technology<br>Rensselaer<br>Rice | 201 | 39.65 | 37.81 | 1 | 1.61 |
| Coeducational liberal arts colleges | Antioch<br>Carleton<br>Grinnell<br>Oberlin<br>Pomona<br>Reed<br>Swarthmore | 162 | 39.64 | 59.26 | 1 | 29.63*** |

*** $p < .001$.

compared with other institutions, turned out to have a marked positive effect ($p < .001$) on the student's PhD aspiration.

The difference between the effects of these types of institutions is considerable when it is realized that, even though the Northeastern men's colleges have a slightly higher *expected* rate of PhD aspirants than the coeducational liberal arts colleges (41% versus 40%), the *actual* rate of PhD aspirants in the coeducational colleges is considerably higher (59% versus 34%). Why these types of institutions should have such different effects is difficult to determine. Both groups are composed of institutions which are highly affluent and which have outstanding reputations. It could be that the climates at the Northeastern men's colleges encourage the student to pursue careers in business and the professions rather than in teaching or research. The coeducational liberal arts colleges, in contrast, may communicate to their students a greater bias toward scientific or academic fields. These same liberal arts colleges, it should be noted, tend to combine several of the college characteristics which are positively related to PhD aspiration: small student bodies with about equal numbers of males and females, and relatively low Conventional Orientations. Just how these college characteristics interact to affect student motivation is, of course, unknown. In any case, a more intensive study of the environments at these institutions would seem to be indicated.

Separate checks on the expected and actual PhD aspiration rates of male and female students again did not indicate that these college effects interact with sex. In the coeducational liberal arts colleges, the actual PhD aspiration rates for both males and females were 1.5 times greater than their expected rates.

## DISCUSSION

In any quasi-experimental work like the present study, there remains the possibility that the significant results are due to uncontrolled input variables, rather than to differential influences of the institutions. However, since a variety of input variables were controlled using a large sample of subjects, and since adjusted PhD aspiration scores were found to be independent of the 20 personality scales, it is unlikely that the marked differences found between expected and actual PhD aspiration rates are not in part the result of different environmental influences.

The design used in this project is probably suitable for the study of college effects on a variety of other student behaviors. However, in cases where (*a*) it would be feasible to regard all input variables as continuous variables, and (*b*) interactions among predictors were not a major source of variance, there is probably something to be gained in terms of efficiency by using multiple regression instead of the actuarial table.

It would seem from the results of this study that the effects of college characteristics on student motivation to seek the PhD are exceedingly complex. None of the college variables appeared to have simple additive effects, and the effect of at least one (the Social Orientation) was clearly nonmonotonic. The fact that a priori groups of colleges (i.e., Northeastern men's colleges and coeducational liberal arts colleges) showed the most pronounced differential effects naturally raises the question as to whether it would eventually be more fruitful to study combinations or configurations of college variables.

The difficulty encountered in interpreting these significant college effects indicates that a more intensive study of student-college environment interactions is necessary. The gross college variables used in this study (i.e., the EAT and the factored dimensions) should continue to be useful in identifying differential college effects, but a more meaningful interpretation of such effects will depend ultimately on our ability to describe the college environment in terms of the experiences of the individual student.

## REFERENCES

ASTIN, A. W. A re-examination of college productivity. *J. educ. Psychol.*, 1961, **52**, 173–178.

ASTIN, A. W. An empirical characterization of higher educational institutions. *J. educ. Psychol.*, 1962, **53**, 224–235. (a)

ASTIN, A. W. Influences on the student's motivation to seek advanced training: Another look. *J. educ. Psychol.*, 1962, **53**, 303–309. (b)

ASTIN, A. W. "Productivity" of undergraduate institutions. *Science*, 1962, **136**, 129–135. (c)

ASTIN, A. W., & HOLLAND, J. L. The Environmental Assessment Technique: A way to measure college environments. *J. educ. Psychol.*, 1961, **52**, 308–316.

ASTIN, A. W., & HOLLAND, J. L. The distribution of "wealth" in higher education. *Coll. & U.*, 1962, **37**, 113–125.

GOUGH, H. G. *Manual for the California Psychological Inventory*. Palo Alto: Consulting Psychologists Press, 1957.

HEIST, P. A. Diversity in college student characteristics. *J. educ. Sociol.*, 1960, **33**, 279–291.

HOLLAND, J. L. Determinants of college choice. *Coll. & U.*, 1959, **35**, 11–28.

SEIGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

STERN, G. G., STEIN, M. I., & BLOOM, B. S. *Methods in personality assessment*. Glencoe: Free Press, 1956.

(Received August 6, 1962)

*Journal of Educational Psychology*
1963, Vol. 54, No. 2, 73–84

# COGNITIVE VERSUS AFFECTIVE FACTORS IN THE LEARNING AND RETENTION OF CONTROVERSIAL MATERIAL[1]

DONALD FITZGERALD
*University of Alberta*
AND DAVID P. AUSUBEL
*Bureau of Educational Research, University of Illinois*

The 264 Illinois high school students studied a passage presenting a Southern interpretation of the Civil War. When general knowledge about the Civil War period was held constant in an analysis of covariance design, the differential effect of attitudinal bias on retention was eliminated. This finding suggests that it is the lack of appropriate subsuming concepts in attitude structure, rather than the direct effect of affective factors on retention processes, that renders "other side" arguments more susceptible to forgetting. Further substantiating the greater relative importance of cognitive factors were the facts that the more knowledgeable Ss learned more than the less knowledgeable Ss. and that an organizing passage facilitated retention.

The phenomena of selective learning and selective forgetting of material involving ego threat has been consistently interpreted in terms of affective mechanisms operating directly on the learning and forgetting processes (Edwards, 1941; Gustafson, 1957; Taft, 1954). Implicit in the interpretive use of affective mechanisms is a denial of the importance of cognitive factors in selective learning and retention. However, despite widespread use of the notion of intrinsic motivational factors in explaining these phenomena, no research to date has unequivocally established the operation of "defense" or "repression" mechanisms in the learning and retention of meaningful material which conflicts with the learner's prevailing attitude structure.

On the other hand, the dependence of learning and retention on cognitive variables has been clearly established (Ausubel & Blake, 1958; Ausubel & Fitzgerald, 1962). The learning and retention of meaningful noncontroversial material have been shown to be directly related to the learner's subject-matter sophistication (Ausubel & Blake, 1958; Ausubel & Fitzgerald, 1962). When existing cognitive structure is generally devoid of subsuming concepts relevant to the learning task, the novel material is poorly learned and rapidly forgotten. This line of research suggests a more parsimonious interpretation of "selective forgetting" than the concepts of defense and repression. In learning "other side" arguments, the conceptual schema constituting the cognitive dimension of attitudes is usually devoid of relevant subsuming ideas to which the new material can be functionally related. The material, therefore, cannot be readily anchored to cognitive structure, competes with existing meanings, and is consequently ambiguous and subject to rapid forgetting. In the case of positive attitudinal bias, it seems reasonable to suppose that the cognitive dimension of attitude structure contains more relevant and appropriate subsuming concepts than in the case of negative bias. Hence the material can be readily anchored to cognitive structure, need not compete with existing meanings, and is therefore less ambiguous and less subject to forgetting. Attitudinal bias might conceivably facilitate or inhibit the *learning* of controversial material by generally raising or lowering levels of effort and concentration. However, once material is learned, it seems more reasonable to attribute differential rates of forgetting to cognitive variables.

---

[1] This paper is based on a dissertation of the same title by the principal author, submitted to the Graduate College of the University of Illinois, in partial fulfillment of the requirements for a PhD in educational psychology.

In order to differentiate experimentally the respective effects of the cognitive and affective components of attitude on learning and retention in this research, attitude structure is analyzed in terms of the direction and intensity of attitudinal bias (i.e., the affective dimension) and the extensiveness and clarity of the related conceptual schema (i.e., the cognitive dimension). An attitude, therefore, is here considered a structure of ideas which involves organization around a conceptual nucleus and which has affective properties (Peak, 1955). That is to say, an attitude involves both affective and cognitive components. Any meaningful concept or conceptual system which an individual internalizes will be related in some way, along a negative-positive continuum, to his ego structure. What we are here calling an attitude is a conceptual nucleus whose evaluative dimension is either highly positive or highly negative from an affective standpoint. The term "cognitive structure" as used in this paper is synonymous with the cognitive dimension of attitude. In contrast to this cognitive component, the phrase "affective components" refers to the complex of feelings that a particular organization of cognition or perception induces in the subject.

Using this affective-cognitive analysis of attitudes, the main purpose of conducting the experiment reported here was to re-examine the effect of attitudes on the learning and retention of controversial material in an attempt to determine the relative contributions of cognitive and affective variables to the learning-forgetting process. Concurrent with this analysis, the efficacy of a brief introductory passage (an "organizer") in facilitating the learning and retention of controversial materials was tested.

In the present study the following hypotheses were tested: (*a*) The clarity and stability of cognitive structure is positively related to the learning and retention of controversial material. (*b*) Positive attitudinal bias facilities, and negative attitudinal bias inhibits, the learning of controversial material. (*c*) Attitudinal bias has no effect on the *retention* of controversial material. (*d*) The learning and retention of controversial material is proactively facilitated by the introduction of a relevant comparative organizer.

## METHOD

### *Subjects and Learning Passage*

Since one purpose of this study was to demonstrate the feasibility of controlled classroom experimentation in this area, 264 junior students of an Illinois high school [2] were used as subjects, and the learning material was related to a traditionally presented unit in United States history.

The experiment was conducted separately in each of 16 American history sections, as one unit in a sequence dealing with the causes of the Civil War. The learning passage consisted of a 2,900-word Southern interpretation of the causes of the Civil War. The material presented a general theme of Northern aggression and acts of violence against the South and, in particular, condemned the Lincoln administration. The regular class unit preceding the experiment considered the same topic as the learning passage but from a Northern point of view. The Southern interpretation was unfamiliar to the subjects and was selected because it conflicted with the culturally biased interpretation of this historical period to which Northern students are continually exposed. By using material which interpreted known facts in an *unfamiliar* controversial way, it was possible both to estimate the clarity and extensiveness of the cognitive dimension of attitude structure (by a pretest covering factual knowledge of the period) and yet to retain the unfamiliarity of the learning passage.

---

[2] The cooperation of the Social Studies Department of Champaign Senior High School is gratefully acknowledged.

## *Test on the Learning Passage*

Knowledge of the Civil War passage was tested by a 34-item multiple-choice test (five responses) with a split-half reliability of .77. The test was used for measuring both learning and retention. Test questions generally covered points of dispute between Northern and Southern interpretations, the five responses being ordered along a continuum from an appropriate Northern response to a response most appropriate in terms of the learning passage. To insure that subjects responded in terms of the learning passage, the test directions repeated the instruction that the response selected should be "the *one* that seems *most* consistent with the ideas and point of view presented in the passage." Test questions were selected by an item analysis from a larger population of items.

Both learning and retention scores on the test showed a satisfactory range of variability and were distributed normally. It was found on item analysis that eight responses could be given a secondary weighting. Correct primary responses were given a weight of 2, and correct secondary responses were given a weight of 1 point, making a maximum score of 68. Since the test was intended as a power test, no time limit was imposed.

## *Attitude Inventory*

A 20-item attitude inventory was prepared to measure the direction and intensity of attitudinal bias toward Southern actions during the era preceding the Civil War. The inventory consisted of controversial statements directly related to the content of the learning passage. The 20 items were selected from a larger population of items on the basis of a group of expert ratings (senior undergraduate students majoring in social studies) of the statements as indicative of either strong Northern or Southern bias. Only items at the extremes of the judgmental scale which manifested little or no variability in the ratings received were chosen. Approximately half the items selected were indicative of the Northern bias, and half were indicative of Southern bias.

Subjects were asked to rate the items in terms of their own agreement, disagreement, strong disagreement, strong agreement, or uncertainty regarding the sentiments expressed in the various statements. The responses were weighted from one to five (five indicating strong Northern bias), and a total attitude score was determined by summing the item scores and dividing by three (for purposes of computation).

Apart from their self-evident face validity as a measure of Northern-Southern bias, an empirical check, carried out with an independent group of subjects, showed that a correlation of .54 ($p < .001$) existed between attitudinal bias scores and intensity of acceptance or rejection of the passage. The group of students used in validating the inventory was a representative sample of junior year students who could not be used in the main experiment because of a schedule conflict. The mean attitude score of the group responding to the passage in a neutral manner was 22.

The attitude scores showed a satisfactory degree of variability and were normally distributed. Since the mean attitude score of subjects in the middle-third of the subgroup distributions was not significantly different from the established neutral point of 22, the upper-third, middle-third, and lower-third attitude groups were described, respectively, as negatively biased, neutral, and positively biased toward the learning passage.

## *Test of Verbal Reasoning*

The Verbal Reasoning test (Form A) of the Differential Aptitude Test (DAT) battery was used as a measure of the subject's ability to abstract, to generalize, and to understand verbal concepts. Approximately 75% of the experimental population had test scores on file. The DAT Verbal Reasoning test was administered to the remaining 25% immediately after the experimental sequence.

## *Introductory Passages*

Two types of introductory passages were used in this experiment, each about 450 words in length. The experimental group studied a *comparative organizer* which explicitly pointed out, at a high level of abstraction, generality, and inclusiveness, the principal differences between the Northern and Southern positions on the causes of the Civil War. The ideas in the comparative organizer were presented at a much higher level of generality and inclusiveness than the learning passage itself, and were deliberately constructed to increase discriminability between the two sets of concepts. The *control introduction* discussed the *history* of Civil War interpretations and the possibility of the same events being interpreted in many different ways depending upon the perspective of the historian, but it contained no substantive material about these interpretations. This control introduction served two important functions. First, it was methodologically important to provide a control introduction in order that any obtained differences in learning and retention outcomes between the experimental and control groups could be attributed to the particular nature of the organizer (i.e., to its enhancing effects on discriminability) rather than to its presence per se. Second, an organizer could conceivably facilitate learning and retention of controversial materials by reducing the negative bias that induces selective learning. Since one principal purpose of this study was to separate the effects of cognitive and affective factors on the learning and retention of controversial material, it was necessary to provide a control introduction that equated the two groups on "affective set," leaving the cognitive effects of the organizer "free" as an independent variable.

Neither introduction provided any information that could be used in answering the test questions on the learning passage.

### *Procedure*

The experimental sequence began immediately following the completion of the unit of study on the events leading up to the Civil War. On the day preceding Session 1, the subjects were informed that some people from the Bureau of Educational Research would be working with them on the pre-Civil-War period. Every attempt was made to have the teachers feel that the material to be studied was relevant to their purposes. It was suggested to the teachers that the results of the pretest on the pre-Civil-War period could be used for grading purposes if they so wished, and that the scores on the pretest as well as on the retention test would be made available to them.

*Session 1.* During Session 1, all subjects took the attitude inventory (8 minutes) and the pretest, and then (after assignment to a treatment group) studied either the organizer or the control introduction for 8 minutes. Membership in a treatment group was determined by random assignment. The population of each of the four treatment groups (i.e., experimental learning, experimental retention, control learning, and control retention) was also stratified by sex so that the proportion of males to females would be the same in each group. This was necessary because of the possibility that female subjects might possess higher DAT scores. It was possible to administer both treatments simultaneously because they consisted of identical appearing introductory passages (with identical sets of directions) differing only in content.

To control for the effects of different teacher, test administrator, and classroom climate variables in the 16 sections, the subjects *within each* section were equally divided among the four treatment groups. Since analysis of the data showed that homogeneity of variance prevailed, both on an interclass and intergroup basis for the Civil War pretest, attitude inventory, and learning and retention tests it was considered justifiable to treat the scores of the four treatment groups on each of these instruments as random samples from the same population.

*Session 2.* The day after Session 1, all subjects read and studied for 28 minutes the learning passage presenting the Southern point of view. Immediately after studying the passage, the experimental and control "learning" groups took the 34-item test on the passage. At the same time the experimental and control "retention" groups took an irrelevant test on Christianity in order to equate the learning and retention groups on procedure. (The topic of Christianity was chosen as maximally dissimilar to the learning passage, and therefore least likely to induce retroactive interference.)

*Session 3.* Seven days after Session 2 all subjects took the test on the passage as a test of retention. The learning groups' scores on this test were not considered in the retention analysis.

## RESULTS AND DISCUSSION

### *Effect of the Organizer on Learning and Retention*

As hypothesized, the organizer treatment was significantly effective in facili-

TABLE 1

ANALYSIS OF VARIANCE OF PRETEST LEVELS × TREATMENT × RETENTION

| Source | *df* | *SS* | *MS* | *F* |
|---|---|---|---|---|
| Treatment | 1 | 417 | 417 | 5.79** |
| Pretest levels | 2 | 12890 | 6445 | 89.5*** |
| Retention | 1 | 4483 | 4483 | 62.3*** |
| Treatment × pretest | 2 | 293 | 147 | 2.04 |
| Treatment × retention | 1 | 1 | 1 | |
| Pretest × retention | 2 | 405 | 203 | 2.82* |
| Treatment × pretest × retention | 2 | 302 | 151 | 2.09 |
| Error | 252 | 18082 | 72 | |
| Total | 263 | 36873 | | |

Note.—Null hypotheses concerning homogeneity of variance accepted.
* $p < .10$.
** $p < .05$.
*** $p < .01$.

tating learning and retention of controversial material ($F = 5.79$, $p < .05$; see Table 1). Moreover the beneficial effect of the organizer was further demonstrated by the significant treatment effect that emerged (*a*) when an analysis of covariance of attitude levels × treatment × retention was performed with DAT held constant ($F = 6.15$, $p < .01$; see Table 2), and (*b*) when an analysis of covariance of pretest levels × treatment × retention, holding DAT constant, was performed ($F = 6.98$, $p < .01$; see Table 3). However, when an analysis of covariance of attitude levels × treatment × retention was carried out with *pretest* held constant, the treatment effect was not significant (see Table 4). Hence it can be reasonably inferred that much of the positive treatment effect evident in the preceding two analyses could be attributed to the interaction (in a nonstatistical sense) between the organizer and what we have called the cognitive dimension of attitude structure.

Although there was a tendency for the experimental group to be superior on learning, most of the benefit derived from the organizer was manifested in relation to retention rather than to learning. When subgroup means of *learning* were corrected for their dependence on DAT, there were no significant differences between treatment pairs of upper-third, middle-third, and lower-third or combined pretest levels (means are given in Table 5). However, comparable corrected retention means indicated a significant difference between the treatment means of the combined levels ($t = 2.31$, $p = .03$; see Table 5 for corrected means), as well as a significant difference between the treatment means in the upper-third pretest level ($t = 2.88$;

TABLE 2

ANALYSIS OF COVARIANCE OF ATTITUDE LEVELS × TREATMENT × RETENTION HOLDING DAT CONSTANT

| Source | *df* | $SS_y$ | *SP* | $SS_x$ | Adjusted | | | *F* |
|---|---|---|---|---|---|---|---|---|
| | | | | | $SS'_y$ | *df'* | *MS'* | |
| Treatment | 1 | 417 | −157 | 59 | 492 | 1 | 492 | 6.15** |
| Attitude levels | 2 | 671 | 3732 | 23192 | 167 | 2 | 84 | 1.05 |
| Retention | 1 | 4483 | 590 | 77 | 4216 | 1 | 1216 | 15.2** |
| Treatment × attitude | 2 | 393 | 241 | 151 | 291 | 2 | 146 | 1.83 |
| Treatment × retention | 1 | 1 | −6 | 1970 | 106 | 1 | 106 | 1.33 |
| Attitude × retention | 2 | 239 | 110 | 4074 | 399 | 2 | 200 | 2.50* |
| Treatment × bias × retention | 2 | 339 | 636 | 1340 | 118 | 2 | 59 | |
| Error | 252 | 30330 | 45282 | 197778 | 19962 | 251 | 80 | |
| Total | 263 | 36873 | 50428 | 228641 | 25751 | 262 | | |

Note.—Null hypotheses concerning homogeneity of variance and regression accepted.
* $p < .10$.
** $p < .01$.

## TABLE 3

ANALYSIS OF COVARIANCE OF PRETEST LEVELS × TREATMENT × RETENTION HOLDING DAT CONSTANT

| Source | $df$ | $SS_y$ | $SP$ | $SS_x$ | Adjusted $SS'_y$ | Adjusted $df'$ | Adjusted $MS'$ | $F$ |
|---|---|---|---|---|---|---|---|---|
| Treatment | 1 | 417 | −44 | 247 | 433 | 1 | 433 | 6.98** |
| Pretest levels | 2 | 12890 | 30278 | 79882 | 11568 | 2 | 5784 | 93.3** |
| Retention | 1 | 4483 | 680 | 103 | 4302 | 1 | 4302 | 69.4** |
| Treatment × pretest | 2 | 293 | 347 | 317 | 207 | 2 | 104 | 1.68 |
| Treatment × retention | 1 | 1 | −5 | 1587 | 30 | 1 | 30 | |
| Pretest × retention | 2 | 405 | −172 | 256 | 455 | 2 | 228 | 3.68* |
| Pretest × treatment × retention | 2 | 302 | 654 | 1678 | 158 | 2 | 79 | 1.26 |
| Error | 252 | 18082 | 18970 | 144242 | 15587 | 251 | 62 | |
| Total | 263 | 36873 | 50708 | 228312 | 25611 | 262 | | |

Note.—Null hypotheses concerning homogeneity of variance and regression accepted.
* $p < .05$.
** $p < .01$.

## TABLE 4

ANALYSIS OF COVARIANCE OF ATTITUDE INVENTORY LEVEL × TREATMENT × RETENTION HOLDING PRETEST CONSTANT

| Source | $df$ | $SS_y$ | $SP$ | $SS_x$ | Adjusted $SS'_y$ | Adjusted $df'$ | Adjusted $MS'$ | $F$ |
|---|---|---|---|---|---|---|---|---|
| Treatment | 1 | 417 | 268 | 172 | 118 | 1 | 118 | 1.74 |
| Attitude inventory level | 2 | 671 | 1237 | 2283 | 72 | 2 | 36 | |
| Retention | 1 | 4483 | 524 | 61 | 3745 | 1 | 3745 | 55.1* |
| Treatment × attitude inventory | 2 | 393 | 260 | 273 | 159 | 2 | 80 | 1.18 |
| Treatment × retention | 1 | 1 | −1 | 34 | 20 | 1 | 20 | |
| Attitude inventory × retention | 2 | 239 | 73 | 114 | 193 | 2 | 97 | 1.43 |
| Treatment × retention × attitude inventory | 2 | 339 | 299 | 375 | 103 | 2 | 52 | |
| Error | 252 | 30330 | 18357 | 25261 | 16990 | 251 | 68 | |
| Total | 263 | 36873 | 21017 | 28573 | | | | |

Note.—Null hypotheses concerning homogeneity of variance and regression accepted.
* $p < .01$.

## TABLE 5

MEANS OF PRETEST LEVELS ON LEARNING AND RETENTION CORRECTED FOR DAT

| Pretest level | Learning $M$ Experimental | Learning $M$ Control | Learning $M$ Total | Retention $M$ Experimental | Retention $M$ Control | Retention $M$ Total |
|---|---|---|---|---|---|---|
| Upper-third | 36.7 | 34.7 | 35.7 | 27.9 | 21.1 | 24.5 |
| Middle-third | 27.0 | 22.6 | 24.8 | 16.9 | 15.2 | 16.1 |
| Lower-third | 21.1 | 21.9 | 21.5 | 16.9 | 16.2 | 16.6 |
| Combined | 28.3 | 26.3 | 27.3 | 20.6 | 17.5 | 19.1 |

Note.—$N$ for each subgroup is 22.

$p = .007$). This latter result is in contrast to the findings of a previous study in which Ausubel and Fitzgerald (1961) found that organizers differentially benefit subjects with a limited and diffuse cognitive structure. The discrepancy can probably be explained by the nature of the learning tasks involved in the two studies.

In the Ausubel and Fitzgerald (1961) study, the learning task dealt with material which, although unfamiliar in terminology, was directly relatable and in many ways similar to concepts already established in cognitive structure. The learning passage dealt with the principles of Mahayana Buddhism, and the organizer presented Buddhism as a categorical variant of Christianity. The comparative organizer, in other words, presented an organized structure of Buddhist concepts in terms of Christianity, thereby clarifying and reinforcing the main principles of Christianity as well as introducing Buddhism. The organizer could therefore be expected to benefit most those subjects whose knowledge of Christianity was unclear, by first clarifying Christianity concepts before differentiating them from Buddhism. Subjects with an extensive and clear organization of Christianity, on the other hand, could be expected to formulate their own differentiations within the learning task itself, and would thus derive little benefit either from the presented contrasts or from the discussion of Christianity given in the organizer. In the present study, however, the ideas presented in the learning passage were concepts which directly conflicted with existing ideas in cognitive structure rather than constituting variations of existing ideas. Whereas the learning task in the previous study tended to reinforce the existing structure of knowledge, the present study was designed to induce cognitive dissonance and affective stress. In the latter circumstances the data clearly indicate that a prerequisite condition for deriving benefit from the organizer is a clear stable cognitive structure of ideas. In a recent study in which the learning task (endocrinology of pubescence) was not directly relatable to general background knowledge in endocrinology, there was a similar tendency for the organizer to benefit differentially those subjects with the higher general endocrinology scores (Ausubel & Fitzgerald, 1962).

### *Effect of Cognitive Variables on Learning and Retention*

The data clearly established the general dependence of learning and retention on pretest scores. In a pretest levels × treatment × retention covariance analysis holding DAT constant, the significant effect of pretest levels was readily apparent ($F = 93.3$, $p < .01$; see Table 3). However, a significant first-order interaction between pretest levels and retention indicated that the effects of the independent variable differed from learning to retention. From Table 5 it can be seen that this interaction is largely a reflection of the fall off from learning to retention in the upper-third pretest level (difference between means of 11.2) in contrast to the much smaller difference in the lower-third pretest level (difference between means of 4.9). Apparently the tendency for the more knowledgeable subjects to manifest a relatively higher rate of forgetting than the less knowledgeable subjects is a function of the greater ideational incompatibility (cognitive dissonance) of the new material with established knowledge in the former instance. Thus, although a clear and stable cognitive structure enables the subjects to learn antithetical new material more readily, this material, once learned, is more rapidly reduced during the retention interval to the original, established propositions in cognitive structure than when existing ideational structure is less clear and stable. Supporting evidence of the long-term inhibitory effect of cognitive dissonance on retention is seen in the significant inverse relationship of .30 ($p < .01$) between degree of pro-Southern attitude and pretest scores: ideational conflict between existing pro-Southern bias in the individual, on the one hand, and the general pro-Northern cultural bias in the Northern community, on the other, tends to render knowledge about the Civil War period less stable than when the individual shares the cultural bias. The higher rate of forgetting in the more knowledgeable group of subjects also explains in part why the organizer has more of an effect on retention than on learning, and why it differentially benefits the subjects with higher pretest scores.

Further evidence of the dependence of learning and retention on pretest scores is provided by simple and partial correlations. Simple correlations between pretest and learning scores ($r = .67$, $p < .01$) and between pretest and retention scores ($r = .59$, $p < .01$) still remained significant when corrected for their dependence on DAT scores. The partial correlation between pretest and learning with DAT eliminated ($r = .47$, $p = .01$) was not significantly different from a similar partial correlation between pretest and retention ($r = .41$, $p < .01$). A similar analysis of DAT by learning and retention indicated a correlation of .68 ($p < .01$) between DAT and learning which was reduced to .49 ($p < .01$) when pretest was partialed out, and a correlation of .59 ($p < .01$) between DAT and retention which was reduced to .22 (*ns*) when the effect of the pretest was eliminated (see Table 6). Therefore it seems that although the clarity and stability of cognitive structure are significant variables in both the learning and retention of controversial materials, verbal reasoning ability is functionally related only to learning.

TABLE 6

PARTIAL CORRELATIONS OF PRETEST AND DAT SCORES WITH LEARNING AND RETENTION SCORES

| Partial correlation | Test | |
|---|---|---|
| | Learning | Retention |
| Pretest [a] | .47* | .41* |
| DAT [b] | .49* | .22 |

[a] DAT eliminated.
[b] Pretest eliminated.
* $p < .05$.

## *Effects of Attitudes and Attitudinal Bias on Learning and Retention*

It has been suggested that the facilitating effect of advance organizers is primarily due to their interaction with the cognitive dimension of attitude structure. The main thesis of this paper is that the cognitive dimension of attitude structure is the major variable in the learning and retention of controversial materials. The intent is not to minimize the obvious significance of *extrinsic* motivational factors in a subject's *general* approach to learning and subsequent recall of the learned materials. What is suggested is that the general interpretation often found in the psychological literature that *intrinsic* motivational or affective factors (e.g., attitudinal bias) account for the phenomenon of selective forgetting in relation to learning tasks involving only a moderate amount of ego threat, is untenable. In order to explore this thesis further in relation to the present study, it is first necessary ($a$) to determine the comparability of the present research to previous studies investigating the phenomenon of selective recall, and

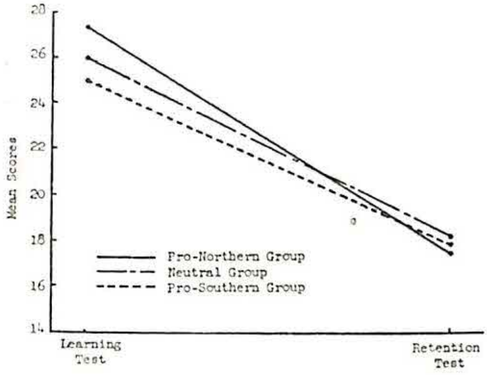


FIG. 1. Mean learning and retention scores for different attitude inventory levels with no statistical control (control group only).

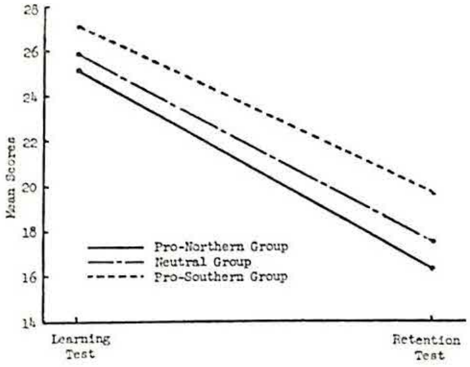


FIG. 3. Mean learning and retention scores for different attitude inventory levels, corrected for pretest (control group only).

(*b*) to clearly distinguish between the effects of cognitive and affective factors on the learning and forgetting processes in order to establish their relative importance.

*Comparability of the Present Study to Previous Research.* No hypotheses were ventured in this study regarding the effect of attitudes (as differentiated from attitudinal bias) on learning and retention, since the nature of this effect has been clearly shown in previous research (Edwards, 1941; Levine & Murphy, 1943; Taft, 1954; Watson & Hartmann, 1939). Nevertheless, an important aspect of the present study is to ascertain the consistency of our general findings with previously demonstrated attitude effects. (Means of the attitude level groups on learning and retention, uncorrected for either pretest or DAT, are shown in Figure 1.)

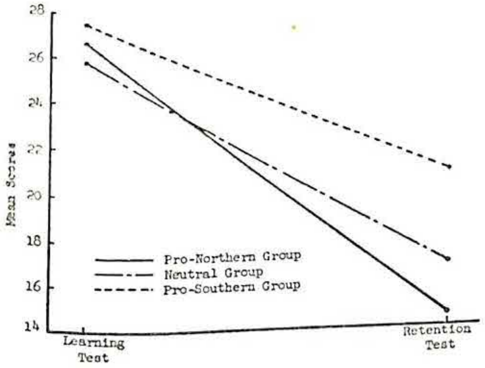


FIG. 2. Mean learning and retention scores for different attitude inventory levels, corrected for DAT (control group only).

The principal attitude effect studied in the learning of controversial materials is the phenomenon of differential rates of forgetting for subjects with negative and positive attitudes, respectively. A general indication of the comparability of the findings of the present study to previous research findings can be obtained by comparing the percentages of material forgotten by negative and positive attitude groups in the study reported herein to corresponding percentages in those previous studies in which both a test of learning and a test of retention were used. These percentages are given in Table 7, ordered in terms of the reten-

TABLE 7

TABLE OF DIFFERENTIAL RATES OF FORGETTING EXPRESSED IN PERCENTAGES AND DF INDEX FOR FOUR STUDIES

| Attitude level | Percentage forgotten | | | |
|---|---|---|---|---|
| | Taft | Present [a] | Levine [b] | Edwards |
| Negative | 42 | 36 | 29 | 13 |
| Positive | 37 | 28 | 21 | 8 |
| DF index | 1.14 | 1.29 | 1.38 | 1.63 |
| Time-days | 3 | 7 | 14 | 21 |

[a] Results of present study uncorrected for independent variables.

[b] Levine and Murphy's results for anti-Soviet-Union materials.

tion period (in days) and rendered comparable by a "differential forgetting" (DF) index. This index was computed by dividing the percentage forgotten by the negative attitude group by the percentage forgotten by the positive attitude group. The general consistency of the present research findings with previously reported results is clearly demonstrated.

*Attitude and Attitudinal Bias Variables.* As previously stated, the affective component of attitudes consists of the general complex of feelings that a particular organization of ideas arouses in the subjects. This affective component is synonymous with attitudinal bias. Levels of attitudinal bias are operationally defined, in terms of our statistical model, as attitude inventory scores with the effect of pretest scores (estimate of the *cognitive* dimension of attitudes) held statistically constant. In contrast, attitude levels as such are operationally defined in terms of attitude inventory scores when pretest is uncontrolled. To introduce a control and reduce the error variance in analyzing the effects of attitude levels on learning and retention, DAT was used as a predictor variable.

*Effect of Attitudes on Learning and Retention.* An analysis of covariance of attitude levels × retention (control group only) with DAT held constant, shows a suggestive but not significant attitude × retention interaction ($F = 1.11$; see Table 8). This same effect was found by Edwards (1941), Taft (1954), and Gustafson (1957) and was interpreted by these investigators as "repression" due to affective or motivated forgetting. In none of these three studies, however, was the interaction effect as pronounced as in the present study. Direction of interaction is clearly seen in the table of attitude level means corrected for DAT (Table 9). This trend is entirely consistent with the findings of related research. Moreover, the "differential forgetting" effect is more strongly suggested by the above data (with the aid of statistical controls) than in the work of Taft, Levine and Murphy, or Edwards. The DF index for the data in Table 9 with DAT held constant is 1.92. It seems, therefore, that the differential forgetting effect which Edwards (1941) interpreted as due to a "conflict with [the] ego's desires" (p. 34), and Taft (1954) interpreted as due to "defense" and "repression" (p. 27), and which Gustafson (1957) referred to as due to an "affective factor in retention" (p. 49),

TABLE 8

ANALYSIS OF COVARIANCE[a] OF ATTITUDE INVENTORY LEVELS × RETENTION HOLDING PRETEST CONSTANT

| Source | *df* | $SS_y$ | *SP* | $SS_x$ | Adjusted $SS'_y$ | *df'* | *MS'* | *F* |
|---|---|---|---|---|---|---|---|---|
| Attitude inventory | 2 | 21 | 461 | 10057 | 313 | 2 | 157 | 2.21* |
| Retention | 1 | 2250 | −1193 | 633 | 2817 | 1 | 2817 | 39.7** |
| Attitude inventory × retention | 2 | 45 | −162 | 738 | 158 | 2 | 79 | 1.11 |
| Error | 126 | 14193 | 23084 | 99892 | 8859 | 125 | 71 | |
| Total | 131 | 16509 | 22190 | 111320 | 12086 | | | |

Note.—Null hypotheses concerning homogeneity of variance and regression accepted.
[a] Control group only.
* $p \leq .13$.
** $p < .01$.

TABLE 9

MEAN LEARNING AND RETENTION SCORES FOR DIFFERENT INVENTORY LEVELS CORRECTED FOR DAT[a]

| Attitude inventory level | Control group M | | |
|---|---|---|---|
| | Learning | Retention | Learning-retention difference |
| Pro-Northern | 26.6 | 14.5 | 12.1 |
| Neutral | 25.8 | 16.8 | 9.0 |
| Pro-Southern | 27.5 | 20.9 | 6.6 |

Note.—N for each subgroup is 22.
[a] Control group only.

can be attributed to the attitude variable as defined above. It is possible in the present research design, however, to consider the relative importance of the cognitive and affective components of attitude in causing "selective forgetting." Clearly, in the work cited above, and implicitly in most other studies of this phenomenon, the interpretation has been entirely in terms of "affect."

*Effect of Attitudinal Bias on Learning and Retention.* The hypothesis that positive attitudinal bias facilitates and that negative attitudinal bias inhibits the learning of controversial materials was rejected. The differences were in the predicted direction (see Table 10) but were not statistically significant.

The differential forgetting trend reported above and shown graphically in Figure 2, although not within an acceptable range of statistical significance, is comparable to that found in previous studies involving controversial material in which the results were also suggestive but nonsignificant. The central focus of research interest in this area has been on the attitude × retention interaction. This interaction effect has been at least as clearly shown in the present study as it has been in previous research.

TABLE 10

MEAN LEARNING AND RETENTION SCORES FOR DIFFERENT ATTITUDE LEVELS, CORRECTED FOR PRETEST[a]

| Attitude inventory level | Control group M | | |
|---|---|---|---|
| | Learning | Retention | Learning-retention difference |
| Pro-Northern | 25.1 | 16.4 | 8.7 |
| Neutral | 25.9 | 17.6 | 8.3 |
| Pro-Southern | 27.1 | 19.8 | 7.3 |

Note.—N for each subgroup is 22.
[a] Control group only.

However, when pretest is held constant in an analysis of covariance of attitude inventory levels × retention (control group only), the inventory level × retention interaction is reduced from 1.13 (pretest not controlled) to .08 (pretest controlled; see Table 11). Comparison of the differences between learning and retention means, both before and after pretest is controlled, also clearly establishes that whatever attitude × retention interaction occurs is entirely attributable to the cognitive dimension of attitudes. When inventory levels are uncorrected for pretest, the difference between the learning and retention means for the pro-Northern-attitude group is 12.1 compared to 6.6 for the pro-Southern-attitude group (see Table 9). When, however, the effect of pretest scores is eliminated, the corresponding differences for the negative and positive attitude groups are 8.7 and 7.3 (see Table 10). The elimination of the attitude × retention interaction is readily seen by comparing Figures 2 and 3.

The hypothesis that attitudinal bias has no effect on the retention of controversial materials was therefore supported. The elimination of the attitude × retention interaction when the effect of pretest was controlled indicated that any selective forgetting must be attribbuted to the cognitive and not to the affective dimension of attitude as previously interpreted by Edwards, Taft, Watson and Hartmann, and others.

TABLE 11

ANALYSIS OF COVARIANCE[a] OF LEARNING AND RETENTION HOLDING PRETEST CONSTANT

| Source | $df$ | $SS_y$ | $SP$ | $SS_x$ | Adjusted $SS'_y$ | Adjusted $df'$ | Adjusted $MS'$ | $F$ |
|---|---|---|---|---|---|---|---|---|
| Attitude inventory | 2 | 21 | 119 | 674 | 155 | 2 | 78 | 1.13 |
| Retention | 1 | 2250 | 66 | 2 | 2162 | 1 | 2162 | 31.3* |
| Attitude inventory × retention | 2 | 45 | 38 | 38 | 11 | 2 | 6 | .08 |
| Error | 126 | 14193 | 8360 | 12492 | 8598 | 125 | 69 | |
| Total | 131 | 16509 | 8583 | 13206 | 10931 | 130 | | |

Note.—Null hypothesis concerning homogeneity of variance and regression accepted.
[a] Control group only.
* $p < .01$.

Through lack of adequate control of the cognitive dimension of attitudes these investigators, in our opinion, have mistakenly attributed differential rates of forgetting in positively and negatively biased groups to intrinsic motivational factors acting on the forgetting process. When adequate controls are introduced it becomes readily apparent that it is the lack of appropriate subsuming concepts in attitude structure that renders "other side" arguments more susceptible to rapid forgetting.

## REFERENCES

AUSUBEL, D. P., & BLAKE, E., JR. Proactive inhibition in the forgetting of meaningful school material. *J. educ. Res.*, 1958, **52**, 145–149.

AUSUBEL, D. P., & FITZGERALD, D. The role of discriminability in meaningful verbal learning and retention. *J. educ. Psychol.*, 1961, **52**, 266–274.

AUSUBEL, D. P., & FITZGERALD, D. Organizer, general background, and antecedent learning variables in sequential verbal learning. *J. educ. Psychol.*, 1962, **53**, 243–249.

EDWARDS, A. L. Political frames of reference as a factor influencing recognition. *J. abnorm. soc. Psychol.*, 1941, **36**, 34–50.

GUSTAFSON, LUCILLE. Relationship between ethnic group membership and the retention of selected facts pertaining to American history and culture. *J. educ. Sociol.*, 1957, **31**, 49–56.

LEVINE, J., & MURPHY, G. The learning and forgetting of controversial material. *J. abnorm. soc. Psychol.*, 1943, **38**, 507–517.

PEAK, HELEN. Attitude and motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1955*. Lincoln: Univer. Nebraska Press, 1955. Pp. 149–189.

TAFT, R. Selective recall and memory distortion of favorable and unfavorable material. *J. abnorm. soc. Psychol.*, 1954, **49**, 23–28.

WATSON, W. S., & HARTMANN, G. W. The rigidity of a basic attitudinal frame. *J. abnorm. soc. Psychol.*, 1939, **34**, 314–335.

(Received October 10, 1962)

*Journal of Educational Psychology*
1963, Vol. 54, No. 2, 85–88

# EFFECT OF CERTAIN SETS UPON COMPLEX VERBAL LEARNING[1]

M. C. WITTROCK
*University of California, Los Angeles*

To test the effect of certain sets upon learning, written passages on Buddhism were given to 143 college students. Before they read the passage 3 groups of Ss were given, respectively, a set to compare, contrast, and compare and contrast Buddhism with Christianity. A 4th group of Ss was given a set to understand and to remember the content of the passage on Buddhism. The data were analyzed by t tests for differences between means. Although Group 1 was not significantly different ($p > .05$, 2-tailed) from Group 4, Groups 2 and 3 were significantly different ($p < .05$, 2-tailed) from Group 4 on both the learning and retention tests. It was concluded that learning sets are important variables in some types of complex verbal learning.

In complex verbal learning one important variable is the interaction between the material to be learned and the students' related information. For example, Ausubel and Fitzgerald (1961) found that the learning of the content of Buddhism was enhanced by written introductions which explicitly compared and contrasted Buddhism and Christianity. In Ausubel's introductions certain concepts were made explicit and were directly given to the learners. At least two factors, learning set and specific information, were probably involved in these introductions. It is possible that simple learning sets, which contain no explicit information about the content to be learned, may affect the learning of Buddhism by actively involving the subjects in comparing and contrasting their related information (Christianity to Buddhism).

In this study written instructions were designed to eliminate explicit information about Buddhism but to establish specific learning sets to: (*a*) compare Buddhism and Christianity, (*b*) contrast Buddhism and Christianity, (*c*) compare and contrast Buddhism and Christianity, and (*d*) understand and remember the content of Buddhism. The term "set" was used only in a strict stimulus-response (S-R) sense.[2] It was assumed that the sets produced by the written instructions listed under *a*, *b*, and *c* above, compared with those of *d* above, increase the probability that the subjects overtly or covertly rehearse certain pre-experimentally determined associations (Christianity) and that the occurrence of these responses facilitates the learning of related material (Buddhism).

## METHOD

### *Subjects*

The subjects were the 143 students enrolled in an upper division educational psychology class at the University of California, Los Angeles. There were 10 men and 133 women. All subjects were required to participate in the experiment. Before the experiment began, each subject was told that he would be informed by personal letter of his own test results and of the results of the whole experiment. Each subject was so informed.

---

[1] The author wishes to thank the Western Data Processing Center and P. Twelker for their important roles in the statistical analysis of the data.

[2] For an excellent analysis of set in S-R mediation theory terms, see Goss (1961, especially pp. 264–268).

### *Materials*

The 36-item, multiple-choice Christianity test; the 2,500-word Buddhism learning passage; and the 45-item, multiple-choice Buddhism learning and retention tests were identical to the materials used and described by Ausubel and Fitzgerald (1961). However, between the page of directions and the beginning page of the content of the Buddhism learning passage, a page labeled Introduction was inserted. On this page two sentences conveyed the learning set. The sentences from the similarity treatment appear below.

> As you read this selection, try to note and remember the many similarities which exist between the Buddhist religion and the religions of the Judeo-Christian tradition with which you are familiar. The main purpose of the study is to determine how well you can learn and remember similarities between the Buddhist religion and those of the Judeo-Christian tradition.

The sentences for two of the other treatments differed from the above sentences only in that the word "similarities" was replaced with the word "differences," or with the words "similarities and differences." The fourth treatment contained two sentences with instructions only to "understand and remember" the content of the Buddhism passage.

A separate sheet of paper was also given to each subject. This page asked him to record, after each of the 9 pages of text, the difficulty of following the instructions in the introduction. A scale of one (not difficult) to four (very difficult) was used. The intent here was to insure that the appropriate set would be maintained as each subject read the entire passage.

### *Procedure*

Since Ausubel and Fitzgerald (1961) found that knowledge of Christianity and verbal ability were each related to performance on the Buddhism test, a stratified random sampling technique was used to assign the subjects individually to the experimental treatment groups. The subjects were stratified according to knowledge of Christianity, as measured by the Christianity test, and verbal ability, as measured by the Cooperative English Test (Educational Testing Service, 1950). To use a stratified random sampling technique was more defensible than a matching technique to reduce the error variance because the latter procedure would probably have introduced a sampling error. It would have been most difficult to match individually four groups on two covariables without eliminating many subjects and without biasing the sample.

To reduce further error variance, all four treatments were administered simultaneously in the same room. Each subject worked individually. There was never more than a total of 15 subjects in the room at one time. Each subject was given a copy of the Buddhism passage which had been designed for his treatment group and a copy of the separate page of instructions. He was allowed 25 minutes to read the Buddhism passage. After he returned the passage to the experimenter, he was given Form A of the Buddhism test. Three weeks later each subject returned to take Form B of the Buddhism test. Both tests were given under untimed conditions.

## RESULTS

The split-half reliability coefficients of the Buddhism test are reported to be .79–.80 (Ausubel & Fitzgerald, 1961). The coefficient of equivalence for Form A and Form B was found by Ausubel to be .79 when the two forms were administered with 1 week intervening. In this study the coefficient of equivalence for the same two forms was found to be .70 with 3 weeks between administrations.

Table 1 presents the means and standard deviations for the treatment groups' performance on the Christianity test, the learning test, and the retention test.

Because the technique used in assigning the subjects to the treatments produced groups quite comparable in mean and standard deviation, $t$ tests for groups which have not been individually matched but which are comparable in mean and standard deviation (Garrett, 1958, p. 230) were appropriate to compare the performance of each experimental group with the control group.

Knowledge of Christianity was correlated .32 and .29 with the learning test scores and with the retention test scores, respectively. Verbal ability was correlated .41 and .44 with these same meas-

TABLE 1

MEANS AND STANDARD DEVIATIONS OF THE CHRISTIANITY COVARIABLE AND OF THE BUDDHISM LEARNING AND RETENTION DATA

| Christianity and Buddhism measures | Treatment group | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Similarities (*N* = 39) | | Differences (*N* = 37) | | Similarities and Differences (*N* = 41) | | Understand and Remember (*N* = 37) | |
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Christianity test | 17.10 | 4.68 | 17.54 | 4.41 | 16.98 | 4.74 | 16.42 | 4.44 |
| Buddhism learning test | 21.23 | 4.89 | 21.57 | 5.32 | 22.39 | 5.06 | 19.19 | 5.44 |
| Buddhism 3-week retention test | 17.59 | 5.41 | 18.65 | 5.76 | 18.17 | 5.21 | 15.78 | 5.11 |

Note.—All scores were the total number of items correct out of a possible 45 for the learning and retention tests, and a possible 36 for the Christianity test.

ures. The *t* ratios reported below were all computed by the formula cited above using the correlation coefficients between performance and Christianity, i.e., .32 and .29.

The assumptions of the *t* test were not violated for each of the *t* ratios presented below. Histograms of frequency distributions for each of the groups involved in the *t* tests were prepared by the IBM 7090 computer. After a visual examination of these histograms it was decided that the assumption of normality of distribution was not violated for any of the treatment groups. To test for homogeneity of variance, *F* ratios were found between the most discrepant treatment group variances on the learning test and on the retention test. These *F* ratios were not statistically significant at the .05 level. From Table 1 it can be seen that the standard deviations of the treatment groups are quite comparable to one another.

The *t* between the Similarities group and the Understand and Remember group was 1.80 ($p > .05$, two-tailed) on the learning test, and 1.55 ($p > .05$, two-tailed) on the retention test. The *t* between the Differences group and the Understand and Remember group was 2.02 ($p < .05$, two-tailed) on the learning test, and 2.35 ($p < .05$, two-tailed) on the retention test. The *t* between the Similarities and Differences group and the Understand and Remember group was 2.81 ($p < .01$, two-tailed) on the learning test, and 2.12 ($p < .05$, two-tailed) on the retention test. Thus, the performance of two of the three experimental groups was as predicted. The performance of the Similarities group was not as predicted. However, its performance was in the predicted direction, and it approached statistical significance.

To provide a further check on the reliability of the data, the three experimental groups were combined into one group and a *t* test was computed between this combined group and the Understand and Remember group. The resulting *t* for the learning test was 2.69 ($p < .01$). The *t* for the retention test was 2.45 ($p < .05$). The mean and standard deviation for the combined group were 21.73 and 5.17 for the learning test, and 18.14 and 5.68 for the retention test.

## DISCUSSION

Provided that they continue to operate throughout the learning interval, some types of sets may enhance the learning of connected discourse material. The results support Ausubel and Fitzgerald's (1961) notion that their introductions contain a learning set.

Although the absolute difference between the experimental groups and the control group was only about two points, this gain must be weighed against the minimal amount of time and effort required to establish and maintain the set. More research is needed to establish whether or not the effect obtained in this study is a real and a general one.

## REFERENCES

AUSUBEL, D. P., & FITZGERALD, D. The role of discriminability in meaningful verbal learning and retention. *J. educ. Psychol.*, 1961, 52, 266–274.

EDUCATIONAL TESTING SERVICE, Cooperative Test Division. *Cooperative English Tests: Form S.* Princeton: ETS, 1950.

GARRETT, H. E. *Statistics in psychology and education.* (5th ed.) New York: Longmans, Green, 1958.

GOSS, A. E. Verbal mediating responses and concept formation. *Psychol. Rev.*, 1961, 68, 248–274.

(Received June 22, 1962)

*Journal of Educational Psychology*
1963, Vol. 54, No. 2, 89–93

# RESPONSE MODE IN THE PROGRAMING OF KINETIC MOLECULAR THEORY CONCEPTS[1]

M. C. WITTROCK
*University of California, Los Angeles*

A kinetic molecular theory program was taught to 80 elementary school children. The experimental group learned the completion item program by responding aloud. The control group listened to the same program without overtly responding. An analysis of variance evidenced an interaction between MA and response mode ($p < .05$). A 1-year retention test evidenced a nonsignificant loss from the learning test. It was concluded that under certain conditions the overt response is probably important in determining the results of learning.

After Keislar and McNeil (1962a, 1962b) established that first grade students could learn kinetic molecular theory from a 13-day, 20-minute per day program, they presented the program via individual teaching machines to primary school children. Two different response modes were used: overt, that is, pushing one of three buttons to select a correct choice from alternatives presented; or nonovert, that is, listening to the program. The overt response mode was used in an operant conditioning procedure to condition to stimuli labels and meanings for concepts. No reliable difference was found between the response modes on paper-and-pencil tests and individual interviews which measured retention and transfer of concepts, but which did not involve "button pushes."[2]

For learning which involves conceptual mediation, vocalization of concept labels is a type of overt motor response which is probably more relevant than is the one used by Keislar and McNeil. The present study used the same type of subject matter and type of subjects used by Keislar and McNeil (1962a) but used vocalization as a type of overt response. It was hypothesized that the learning of meanings of words can be facilitated by the conditioning of verbalizations of these words to the auditory and visual stimuli of a program designed to teach concepts. For discussions of the conditioning of word meanings, see Mowrer (1960, pp. 117–162), Osgood (1953, pp. 680–727), and Staats (1961).

## METHOD

### *Kinetic Molecular Theory Program*

The program consisted of images from colored slides made from artists' drawings and of commentary played from a tape recorder. The program was designed to teach kinetic molecular theory, including the definition of molecules, the movement of molecules, various states of matter, evaporation, and condensation.

For the overt response group, the program was written in completion item form. The last word of most of the sentences was to be supplied by the subjects during 5-second pauses in the commentary. The beginning of each pause was indicated by an audible tap of a bell. After the pauses, the taped commentary presented the correct words. The words which the students were required to speak aloud were labels for concepts, such as "air," "molecule," or "gas," or labels for characteristics of the concepts. The program was written to provide cues in the forms of introductions and explanations to elicit the labels.

---

[1] The author wishes to thank E. Keislar and J. McNeil for their wholehearted cooperation in performing the study, and especially for the use of the materials and tests from their earlier research.

[2] Other recent studies, which are briefly reviewed by Krumboltz (1961), also report "no differences" between overt and nonovert response modes. However, there is considerable evidence, which is reviewed in detail by Lumsdaine (1963), that the overt response can facilitate different types of learning.

For the control group, the slides and tapes were identical to the ones used for the experimental group, except that the 5-second pauses and the brief introduction on overtly responding did not occur. For both groups, the program was written in eight consecutive parts, each part 15–20 minutes in length.

After the initial writing of each part, the program was individually administered to first graders in a school different from the one in which the experiment was performed. Six boys and six girls were chosen to represent varying levels of IQ and achievement. All student responses were recorded and used in refining and rewriting the sequence and content of the frames of the program. After individual sessions with four students, the program was revised according to the feedback obtained. This process of presenting and revising was repeated two more times with the other eight students.

### *Subjects*

The 80 first and second grade subjects were divided into two individually matched groups of 40 each according to mental age (MA), and as nearly as possible, according to chronological age (CA), IQ, and sex. The mean and standard deviation for MA of the overt group were 84.3 and 9.7. For the nonovert group these figures were 84.5 and 10.6. For IQ these figures were 119.7 and 13.9 for the overt group, and 119.1 and 14.6 for the covert group.

The subjects of this study were taken from a school district immediately adjacent to School A of the Keislar and McNeil (1962a) study. The subjects from these two studies were comparable in IQ, MA, CA, and in socioeconomic status.

### *Procedure*

The program was presented in a room equipped with 10 tables partitioned from each other. The subjects were run in groups of 10 or less. Each child wore a set of headphones through which he heard the taped commentary. Two microphones attached to two additional tape recorders operated in the room. These two tape recorders were controlled to operate only during the 5-second pauses in the tapes for the overt response groups. Thus, there was a record of the oral responses of 8 of the children. Experimental and control groups were run in ABBA fashion to counterbalance the time of day. The program was completed in 8 consecutive school days. Immediately after the end of the instruction period, the 10-minute individual, standardized interview and the written, multiple-choice group tests described elsewhere (Keislar & McNeil, 1962a) were given.

One year later the written, multiple-choice group test mentioned above was given again to the subjects.

## RESULTS

The error rate obtained from the recordings of the responses of the overt

TABLE 1

SCORES OF THE OVERT AND NONOVERT GROUPS ON THE WRITTEN POSTTESTS

| Group | Posttest | | | | | |
|---|---|---|---|---|---|---|
| | Learning test | | | 1-year retention test | | |
| | *N* | *M* | *SD* | *N* | *M* | *SD* |
| Above median MA | | | | | | |
| Overt | 20 | 27.80 | 5.08 | 17 | 26.88 | 4.86 |
| Nonovert | 20 | 29.70 | 4.65 | 17 | 26.05 | 4.05 |
| Below median MA | | | | | | |
| Overt | 20 | 26.60 | 5.88 | 19 | 24.17 | 6.23 |
| Nonovert | 20 | 23.85 | 6.13 | 19 | 24.25 | 5.86 |
| Combined overt | 40 | 27.20 | 5.52 | 36 | 25.45 | 5.54 |
| Combined nonovert | 40 | 26.78 | 6.16 | 36 | 25.10 | 4.95 |

Note.—Possible score was 38.

TABLE 2

SUMMARY OF THE RANDOMIZED BLOCKS ANALYSIS OF VARIANCE FOR THE WRITTEN LEARNING TEST

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| MA | 1 | 248.52 | 11.15** |
| Response mode | 1 | 3.62 | — |
| Interaction | 1 | 108.10 | 4.85* |
| Blocks | 39 | 40.22 | |
| Within | 37 | 22.28 | |
| Total | 79 | 422.74 | |

*Note.*—$N = 80$.
* $p \leq .05$.
** $p \leq .01$.

group was 26%. An error was recorded when the subjects failed to respond or responded incorrectly.

A summary of the results for the written posttests are presented in Table 1. Homogeneity of variance for the groups reported in Table 1 was assumed after *F* tests produced no statistical reasons ($p > .05$) to discredit the assumptions. Normality of distribution was assumed after histograms of the frequency distributions for the treatment groups were examined.

### *Written Learning Test*

Because the overt and nonovert subjects had been individually matched for MA, a 2 × 2 factorial, randomized block design (Edwards, 1960, pp. 213–215) was an appropriate analysis of variance. The two factors were: (*a*) above and below median MA, and (*b*) overt and nonovert response modes. The blocks were the 40 pairs of matched subjects. The variance due to interaction between response mode and MA was statistically significant ($p < .05$). From Table 1 the overt response mode was more effective with the below median MA subjects than it was with the above median MA subjects. Performance was correlated .40 with MA and .40 with IQ.

### *One-Year-Later Test*

The retention and transfer test evidenced no significant interaction between MA and response mode. The means for the overt and covert response groups showed no statistically significant loss in retention over the 1-year period.

### *Measures of Efficiency*

The running time of the tapes was 120.2 minutes for the nonovert group and 165.0 minutes for the overt group. By dividing the criterion measures presented in Table 1 by the respective running time in minutes of the taped program, measures of efficiency were obtained. The efficiency measures were each multiplied by 100 and are presented in Table 3. As this table shows,

TABLE 3

EFFICIENCY MEASURES FOR THE OVERT AND THE NONOVERT GROUPS ON THE LEARNING TEST

| Subject | *N* | *M* | *SD* | $M_{\text{Overt}} - M_{\text{Nonovert}}$ | *t* |
|---|---|---|---|---|---|
| Above median MA | | | | | |
| Overt group | 20 | 16.85 | 3.08 | −7.86 | 6.96** |
| Nonovert group | 20 | 24.71 | 3.87 | | |
| Below median MA | | | | | |
| Overt group | 20 | 16.12 | 3.56 | −3.72 | 2.20* |
| Nonovert group | 20 | 19.84 | 5.10 | | |
| Combined overt subjects | 40 | 16.48 | 3.35 | −5.80 | 6.67** |
| Combined nonovert subjects | 40 | 22.28 | 5.12 | | |

* $p \leq .05$.
** $p \leq .01$.

the nonovert group was more efficient ($p < .01$, two-tailed) than the overt group.

## DISCUSSION

The interaction between response mode and MA implies that relevant overt responses may enhance learning of difficult concepts by primary school children who are average or nearly average in scholastic ability (IQ below 120), but that relevant overt responses may add nothing to learning of concepts by children with high scholastic ability (IQ above 120). However, the interaction was not statistically significant 1 year later.

After this study was completed Goldbeck and Campbell (1962) reported a related finding. They found an interaction between response mode and difficulty level.[3] The positive results obtained in the present study and those by Goldbeck and Campbell (1962) do not agree with the studies mentioned in the introduction which report no differences between response modes. It would be premature to conclude on the basis of these latter studies that response mode is not an important variable in classroom learning. These no differences findings, as well as the retention test results of the present study, probably evidence a need for more sensitive measuring instruments than are presently available, and for continued research on mode of response with relevant operants.

At face value, the retention test results imply that programed learning which is designed to develop concepts can produce a high percentage of retention. These scores may be high for several spurious reasons: (*a*) practice from the learning test, (*b*) intervening learning, and (*c*) intellectual development by the subjects. The results do not indicate that the children knew the material before the experiment began. On the tests used in the present study, Keislar and McNeil (1962b, p. 107) found that primary school children who had not studied the program scored near chance and definitely below the scores of the groups who had studied the program.

One could argue that the "1-year-later" scores of the present study should have been higher than the learning scores (e.g., Tyler, 1934) because further learning and application of concepts should have followed during the year. However, such findings are rare. The retention test scores appear to be unusually high.

This study supports Mogar (1960) and Keislar and McNeil (1962a, 1962b) who found that the primary school child's conception of physical relationships can be enhanced by instruction. Not only can some primary school children learn, retain, and transfer the rudiments of kinetic molecular theory, but they can learn them very quickly.

## REFERENCES

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1960.

GOLDBECK, R. A., & CAMPBELL, V. N. The effects of response mode and response difficulty on programed learning. *J. educ. Psychol.*, 1962, **53**, 110–118.

KEISLAR, E. R., & MCNEIL, J. D. A comparison of two response modes in an autoinstructional program with children in the primary grades. *J. educ. Psychol.*, 1962, **53**, 127–131. (a)

KEISLAR, E. R., & MCNEIL, J. D. Teaching science and mathematics by autoinstruction in the primary grades: An experimental strategy in curriculum development. In J. E.

---

[3] The overt response groups performed below the covert response groups at the lowest difficulty level but above the covert response groups at the intermediate difficulty level. The subjects were seventh graders and the tasks were independent factual items from social studies. Insofar as difficulty level of scholastic tasks and scholastic ability are inversely related, their results and the results of the present study are in agreement.

Coulson (Ed.), *Programed learning and computer based instruction.* New York: Wiley, 1962. Pp. 99–119. (b)

KRUMBOLTZ, J. D. Meaningful learning and retention: Practice and reinforcement variables. *Rev. educ. Res.,* 1961, 31, 535–546.

LUMSDAINE, A. A. Instruments and media of instruction. In N. L. Gage (Ed.), *The handbook of research on teaching.* Chicago: Rand McNally, 1963. Ch. 12.

MOGAR, MARIANNINA. Children's causal reasoning about natural phenomena. *Child Develpm.,* 1960, 31, 59–65.

MOWRER, O. H. *Learning theory and the symbolic processes.* New York: Wiley, 1960.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford, 1953.

STAATS, A. W. Verbal habit-families, concepts, and the operant conditioning of word classes. *Psychol. Rev.,* 1961, 68, 190–204.

TYLER, R. Some findings from studies in the field of college biology. *Sci. Educ.,* 1934, 18, 133–142.

(Received June 25, 1962)

*Journal of Educational Psychology*
1963, Vol. 54. No. 2, 94–103

# STRUCTURE OF ABILITIES IN A NON-WESTERN CULTURE[1]

GEORGE M. GUTHRIE
*Pennsylvania State University*

A battery of 50 tests including some tests in the Philippine dialect, Tagalog, was administered to 314 Philippine women college students. 22 factors were extracted. Verbal, numerical, rote memory, visualization, fluency, and speed factors were found to be very similar to those found by others. Several factors were found which appeared to be a product of the Philippine educational tradition. Memory and reasoning factors were somewhat ambiguous. The Tagalog verbal factor was quite separate from the English verbal factor. Some implications for further research using more matching tests in the non-Western language were suggested.

Our knowledge of the factorial structure of ability is based almost exclusively on data gathered from English speaking subjects. The degree to which factors found and confirmed in one culture will be found also in markedly different cultures remains unknown. Klineberg (1935) has summarized the evidence that cultural factors make it virtually impossible to evaluate levels of ability across different cultures. For this reason, the conclusion of Porteus and Babcock (1926) that marked differences in intelligence exist between different races has been questioned. The effect of environmental influences has been demonstrated within our own society where subtle cultural factors appear to be responsible for some of the differences between means found on different social groups. But the organization of abilities is another matter, and evidence is needed on the degree to which factorial structure remains constant across widely different cultures.

Among the factors that could produce differences in the factorial structure of abilities are the linguistic system, genetics, enviromental demands, and the mode of life of the subjects. Whorf (1956, pp. 78–83) made rather dramatic speculations about the manner in which different linguistic systems may mold a person's way of thinking. In order to evaluate the differences he postulates, we would have to study subjects whose linguistic system is considerably removed from our own, preferably some system outside of Indo-European languages. The Whorfian hypothesis, intriguing as it is, has been difficult to reduce to verifiable implications. With respect to hereditary influences, Vernon (1951) insists that genetics may account for 75% of the variance in the level of ability, but he does not consider evidence on the role of genetic factors in the structure of abilities. If hereditary factors influence the level, it is possible that they also influence the correlations between samples of various abilities. Environmental demands lead to differing levels of development of abilities but whether specific demands can lead to specialized factors remains uncertain. For instance, do members of a society which leans strongly on authority and tradition develop reasoning abilities which are organized differently? Finally, the pattern of relationships between people may lead to subtle skills of interpersonal perception or memory. In some areas a different language must be used in speaking to a superior than in speaking to a subordinate. Does there emerge in such a society a set of correlated performances related to these role shifts? In short, we know that our tests are to some degree culture bound, but we do not know how general are our factors.

---

[1] These data were collected during a Fulbright appointment at the Philippine Normal College, Manila. The author wishes to acknowledge the help of E. C. Ramirez, President, and his staff. He also wishes to acknowledge the helpful advice of J. W. French, S. Vandenberg, and D. R. Saunders. Finally, the analysis was made possible by D. Laird and the staff of the Computation Center of Pennsylvania State. A paper based on these results was presented at the annual American Psychological Association Convention, New York, September 1961.

Since language differences are so obvious, nonverbal tests have often been used in the expectation that these would provide adequate measures which would be relatively free of cultural factors. The reports of Jahoda (1956) and Scott (1950), however, indicate that, in going to another culture, it is not enough simply to substitute nonverbal for verbal tests. We do not know what we are substituting because there are few reports of factored batteries on other than English speaking subjects.

Nadel (1936, p. 195), reviewing the application of psychological tests in anthropological work, concluded,

> We must admit the existence of different types of intelligence, and of the varied roles which different societies assign to them; we must also take them into account in constructing our tests of intelligence. . . . We should not so much be examining and measuring degrees of intelligence as analyzing different types and qualities of intelligence.

Eighteen years later Vernon (1954) observed,

> We must recognize that intelligence itself is a different entity in each group. . . . Western civilization values highly symbolic reasoning . . . another group's intellect might be most fully extended in a very different field of activity say religious rituals, hunting or farming. It would be necessary then to devise tests within this field, and by means of factor analysis to arrive at the structure or makeup of the Highest Common Factor or g in the group's abilities. But if cultures were not too dissimilar, tests might be found which had good g loadings in both.

Biesheuvel (1949), who has also had experience testing non-Western groups, after surveying factors that could influence test scores, suggested that by means of factor analysis "it should be possible to discover not only the basic constituents of intelligence in any given case, but also its cultural and genetic determinants." This may be too much to hope for. He has developed his thesis with special reference to the study of African ability (Biesheuvel, 1952a, 1952b).

Some of the questions about the generality of primary mental abilities may be answered by administering batteries of tests in English and in some non-Indo-European language to bilingual subjects. Vandenberg (1959) has conducted such a study with Chinese students studying at American universities. He found five factors: spatial, verbal, number, memory, and perceptual speed which closely approximated those of Thurstone. In addition he found an ambiguous reasoning factor and a separate Chinese vocabulary test factor. He concluded that,

> cultural influences play a role in the process leading to the formation of abilities underlying some of the factors, but that at least several potentialities exist in the adult human neurophysiological organization that are independent of one another and, to some extent, independent of the particular kind of cultural, linguistic, and educational background of the subjects tested.

Since the research to be reported was done in the Philippines it should be pointed out that, in addition to the work of Porteus, others who have used ability tests in the south Pacific area include Carreon (1926), Dunlap (1931), and

Mann (1939). Each used tests in English, verbal and nonverbal; the closest approach to factor analysis was Dunlap's use of tetrad difference equations. A more detailed examination of the pattern of correlations between various tests should answer the question whether we can assume the same ability dimensions in another society.

This study is concerned with the structure of abilities in a non-Western population. The purpose is to determine the degree to which the primary mental abilities found in American samples will also appear in subjects with a different genetic, cultural, and linguistic background.

The Philippines present an unusual opportunity to clarify some of the issues on the generality of factors and the role of cultural and biological determinants in the differentiation of abilities. The population is of the Malayan race. They speak a number of Malayo-Polynesian dialects in their homes. Their environmental demands have been quite different. Their family system and patterns of interpersonal relationships depart in many ways from those of the west. At the same time, for the past 50 years they have had English as a language of instruction in the schools. Here, then, is a bilingual population to whom can be administered a standard battery of tests which has been used in American studies. While the level of performance should not be directly compared, the pattern of correlations can be examined to determine the degree to which intellectual activities group in the same manner that they do in this country. From this we may infer how primary are the primary mental abilities.

Several questions influenced the selection of tests. Would the frequently identified abilities emerge again? Would the verbal factor, which accounts for so much variance in American studies, appear as prominently in the Philippines with its less developed education emphasis? Is memory structured in the same way in the Philippines where, with less reliance on writing, memory functions are more emphasized? In view of the perplexity Americans have sometimes experienced with Philippine reasoning processes, are reasoning factors organized there as they are in societies where people appear to reason in more familiar ways? Finally, what is the effect of giving a parallel test in a Philippine dialect?

## METHOD

### *Subjects*

The subjects were women college students in education in their junior year. All were between the ages of 19 and 22. English had been the language of instruction throughout their entire school experience. More than three-quarters of the group came from the provinces of the Tagalog speaking regions near Manila on the island of Luzon. None listed English, Spanish, or Chinese as the language spoken in her home. About two thirds listed Tagalog as her home language while the remainder listed other Philippine dialects, such as Ilocano, Bicolano, or Cebuano, all members of the Malayo-Polynesian language family. This means that all subjects were bilingual, speaking at least English and Tagalog. Many spoke several other Philippine dialects as well. Nearly all subjects appeared to revert to Tagalog or another dialect outside of the classroom. The subjects, therefore, had been raised in a Philippine language environment and had learned English in their school experience. They could take tests in either English or Tagalog.

Of some 350 subjects who were asked to participate in this study, 314 completed all of the tests. The tests were administered in groups as large as 75 in four 2-hour sessions, spaced over several weeks. The author and several Philippine proctors were present at all sessions. In order to insure familiarity with the tasks, somewhat longer explanations and practice sessions were given. All instructions were in English.

### *Development of the Test Battery*

The battery of tests which was finally selected is shown in the order in which it was administered. The factor on which each test may load is suggested. Most of the tests were drawn from French's (1954) kit. Additional tests were prepared in Tagalog, the Manila area dialect, to measure the verbal factor and ideational and verbal fluency. Tests were selected which had proven to be satisfactory measures of the verbal, numerical, spatial, and visualization factors. There were also tests of motor and of perceptual speed; of inductive, deductive, and general reasoning; and of both rote and meaningful memory. Memory functions were given special attention, with the selection of tests based on the work of Kelley (1954).

Other considerations influenced the selection of tests. Since there appears to be a good deal of emphasis upon memorizing in Philippine education and less on problem solving, tests of memory functions were included with both meaningful and rote material, and with visual and auditory presentations. An index of academic achievement was included. Reasoning processes was another area of special interest. In order to control the influence of perceptual and motor speed appropriate tests were included. Similarly, fluency could spuriously raise correlations if not accounted for.

### *Tests*

1. Anagrams in Tagalog—Make Tagalog words from an 11-letter Tagalog word. Word fluency.

2. Topics in Tagalog—Write in Tagalog ideas about a given topic. The test topic is given in Tagalog. Ideational fluency.

3. Thing categories in Tagalog—List things in Tagalog that could have a certain characteristic. Ideational fluency.

4. Suffixes in Tagalog—List words in Tagalog ending with a given set of letters. Word fluency.

5. Following directions—Given a 5 × 5 matrix of letters, find which letter is indicated by increasingly complex directions. Verbal and induction. (Detroit Advanced Intelligence Test.)

6. Analogies—Verbal. (Carnegie Mental Abilities Test.)

7. Number Series I—Pick out the number in a row of seven numbers which does not belong. Induction. (Detroit Advanced Intelligence Test.)

8. Disarranged Sentences—Sentences of from 5 to 12 words are presented with the word order scrambled. The subject is to indicate whether the sentence is true, false, or incomplete. Verbal and several other factors. (Carnegie Mental Ability Tests.)

9. Anagrams—Make words from a 12-letter English word. Word fluency. (Thurstone, 1941.)

10. Number Series II—Select which of five alternative pairs of numbers will continue a series of numbers for which an introductory series of from four to seven have been presented. Induction. (Carnegie Mental Abilities Test.)

11. First and last letters in Tagalog—Words in Tagalog which begin and end with certain given letters. Word fluency.

12. Letter series—Fill in the next letter in a given series. Induction. (Thurstone, 1941.)

13. High Numbers—Draw a circle around each one- or two-digit number which is higher than the number on either the right or the left. Space, Perceptual speed. (Thurstone, 1941.)

14. Spelling—Select the correctly spelled alternative from five. Word fluency, verbal. (Detroit Advanced Intelligence Test.)

15. Vocabulary—Select the word that means the same from four alternatives. Verbal. (French, 1954.)

16. Word Squares—Fill in the missing word in a matrix of words from a list of words. Words are selected to fulfill relationship suggested in the rows and columns. Deduction. (French, 1954.)

17. Same or Opposite—Draw a circle around a word which means the same as or the opposite of the first word in a row of five words. Verbal. (Thurstone, 1941.)

18. Thing categories—Name as many things as possible that fall in a given category. Ideational fluency. (French, 1954.)

19. Writing LACK—Write this word as rapidly as possible for a given period. Speed. (French, 1954.)

20. Suffixes—Give as many words as possible that end in a given suffix. Word fluency. (French, 1954.)

21. Writing Digits—Write the Numbers 1–9 as rapidly as possible. Speed. (French, 1954.)

22. Mathematics Aptitude Test—Solve a series of arithmetic problems and select the right answer among five alternatives. General reasoning. (French, 1954.)

23. Division—Series of division problems with a one-digit divisor. Numerical. (French, 1954.)

24. Letter Groups—Select which four-letter group does not belong in four sets presented. Induction. (French, 1954.)

25. Writing x—write x over a series of dots. Speed. (French, 1954.)

26. Digit Symbol—Write the same symbol under a digit as is found under that digit in a key at the top of the page. Memory. (Follows style of Wechsler.)

27. Reasoning—Draw a conclusion from two given sentences by filling in a missing word. Deduction. (French, 1954.)

28. Addition—Add up to three two-digit numbers. Numerical. (French, 1954.)

29. First Digit Cancellation—Mark out those instances in a row where the first digit of the row appears. Speed of discrimination. (French, 1954.)

30. Subtraction and Multiplication—Alternate rows of subtraction and multiplication problems. Numerical. (French, 1954.)

31. Verbal Classification—Categorize each of six words in one of two lists of four words each. Verbal. (Thurstone, 1938.)

32. Identical Forms—Select one of five forms which is identical with the first form in the row. Perceptual. (Thurstone, 1938.)

33. Paper Form Board—Mark an outlined figure to show exact fit of a group of solid black pieces. Visualization. (French, 1954.)

34. Picture-Number Memory—Memorize the two-digit number that goes with each of 21 simple pictures. Memory. (French, 1954.)

35. Punched Holes—Mark the holes on a piece of paper if the hole had been punched with the paper folded a number of times. Visualization. (French, 1954.)

36. Figure Recognition—Pick out of a group of 60 pictures those 20 which were seen before. Memory. (Thurstone, 1938.)

37. Tagalog-non-Tagalog—The subject indicated whether or not Tagalog was the language spoken in her home.

38. Grade-Point Average—The subject's grade-point average during her junior year in college.

39. Completion—Fill in word to complete sentence with first letter of the word given. Verbal. (Thurstone, 1938.)

40. First Names—Recall first name when presented last name of 20 persons' names. Memory. (French, 1954.)

41. Word-Number Memory—Recall two-digit numbers paired with noun. Memory. (Thurstone, 1938).

42. Digit Span—Digits presented orally in groups of three to nine, immediate recall. Memory.

43. Manila Story Recall—A three-paragraph story was presented orally. Recall as many ideas as possible. Meaningful memory. (After Kelly, 1954.)

44. Flags—Indicate whether same face of flag is shown on pairs of flags. Spatial. (Thurstone, 1938.)

45. Number Code—Learn a new number system on base 20 instead of 10. Numerical. (Thurstone, 1938.)

46. Tagalog Vocabulary—Select synonym for each word from four alternatives. Verbal. (Prepared especially for this research.)

47. Topics—Write as many ideas as possible about a suggested topic. Ideational fluency. (French, 1954.)

48. First and Last Letters—Write as many words as possible beginning and ending with prescribed letters. Word fluency. (French, 1954.)

49. Number-Number Memory—Recall second number when presented first number of a pair. Memory. (Thurstone, 1938.)

50. Paragraph Recall—Fill in missing words in a paragraph that had been presented orally. Meaningful Memory. (Thurstone, 1941.)

In the testing sessions it was found that the subjects tended to give up rather quickly if the items became difficult, but they always completed the test even though they guessed after the first few items. Rarely were items omitted. The Cubes test and the False Premises test, with a two-alternative format, could not be used because the results clustered symmetrically around chance. There was little difficulty with subjects helping each other, but it was difficult to keep them from turning ahead to new tests before the instruction was given. The subjects experienced greatest difficulty understanding tests of spatial abilities. The reasons for this are not too clear, but others report the same difficulty with other subjects in this part of the world. In summary, the subjects were cooperative, but they tended to answer all items, reverting to guessing rather than sustaining effort.

## RESULTS

### *Analysis of Data*

Product-moment correlations were computed between the scores of the tests and the matrix was first analyzed using the principal components solution with unities in the diagonal. Fourteen factors were extracted and rotated using Kaiser's varimax solution. The results are shown in Tables A–C.[2]

Subsequently the data were analyzed using a derived estimate of communalities in the diagonal (Saunders, 1960), and 22 factors extracted by the principal components method were rotated using both the varimax and ratiomax (Saunders, 1962) solutions. The solutions are presented in Tables D–F (see Footnote 2). The extraction of such a large number of factors had the effect of dividing several factors found when only 14 were extracted. The ratiomax solution distributed the variance over the factors more evenly so that later factors accounted for more of the variance and accordingly were somewhat more readily interpreted. The results of the ratiomax solution are presented listing all tests with loadings of at least .25 on each factor.

### *Interpretation*

Of the 22 factors, 16 appear to be interpretable. The other 6 have only one high loading and, at this point, remain ambiguous. The factors are presented listing all tests with loading above .25.

| Factor I | Numerical Facility | Loading |
|---|---|---|
| 30 | Subtraction and Multiplication | .77 |
| 23 | Division | .72 |
| 28 | Addition | .65 |
| 7 | Number Series I | .27 |
| 45 | Number Code | .26 |

This is a frequently found factor which is quite sharply delineated in this population. The simple arithmetic operations appear in a factor separate from other factors which also involve the manipulation of numbers.

| Factor II | Verbal Comprehension | Loading |
|---|---|---|
| 15 | Vocabulary | .68 |
| 17 | Same or Opposite | .66 |
| 39 | Completion | .43 |
| 38 | Grade-Point Average | .35 |
| 14 | Spelling | .28 |
| 8 | Disarranged Sentences | .26 |

This factor is specific to content presented in English. The Grade-Point Average loads higher on this factor than on any other factor. Even though facility in English is lower in this group, the verbal factor is sharply separated from reasoning and fluency factors which also involve words.

| Factor III | Associative (Rote) Memory | Loading |
|---|---|---|
| 34 | Picture-Number Memory | .72 |
| 41 | Word-Number Memory | .69 |
| 40 | First Names | .39 |
| 49 | Number-Number Memory | .35 |

All of the high loadings on this factor involve the association of material that does not have any particular meaningful relationship.

| Factor IV | Visualization | Loading |
|---|---|---|
| 33 | Paper Form Board | .73 |
| 35 | Punched Holes | .66 |
| 16 | Word Squares | .25 |

The two visualization tests are followed by a reasoning test with a low loading. These two factors are often closely related. The differentiation of spatial ability and visualization which has been somewhat uncertain, appears here since the Flags test does not load their factor.

| Factor V | A Factor Specific to Anagrams | Loading |
|---|---|---|
| 1 | Anagrams in Tagalog | .68 |
| 9 | Anagrams | .64 |
| 6 | Analogies | .26 |

When fewer factors were extracted, these three tests appeared in the word fluency factor with high loadings. It would appear that this should be considered a word fluency factor specific to anagrams regardless of the language of the test.

---

[2] Tables A–F have been deposited with the American Documentation Institute. Order Document No. 7464, from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

| Factor VI | Motor Speed | Loading |
|---|---|---|
| 25 | Writing X | .63 |
| 19 | Writing LACK | .60 |
| 21 | Writing Digits | .36 |

This is clearly a motor speed factor on which some of the fluency tests have very low loadings.

| Factor VII | Tagalog-Non-Tagalog | Loading |
|---|---|---|
| 37 | Tagalog-non-Tagalog | .74 |
| 46 | Tagalog Vocabulary | .61 |
| 2 | Topics in Tagalog | .26 |
| 1 | Anagrams in Tagalog | .26 |

The factor is specific to the dialect background of the subjects. The vocabulary test in Tagalog loads this factor and not the general verbal factor. It is relatively specific to vocabulary and shows only a low loading of one of the fluency tests in Tagalog which were included in the battery.

| Factor VIII | Ideational Fluency | Loading |
|---|---|---|
| 47 | Topics | .79 |
| 2 | Topics in Tagalog | .36 |
| 18 | Thing Categories | .30 |
| 3 | Thing categories in Tagalog | .28 |

This factor shows rather clearly that ideational fluency is not specific to the language system in which the subject is asked to work.

| Factor IX | Inductive Reasoning | Loading |
|---|---|---|
| 12 | Letter Series | .59 |
| 16 | Word Squares | .50 |
| 24 | Letter Groups | .29 |

This factor was separated from a larger factor which included the visualization tests when the more extended factor extraction was carried out. In each of the tests the subject has to work out the pattern that is offered and then supply or select the item that completes or continues the pattern.

| Factor X | Reasoning with Number Patterns | Loading |
|---|---|---|
| 5 | Following directions | .61 |
| 7 | Number Series I | .56 |
| 10 | Number Series II | .36 |

This factor is somewhat obscure. It appears to involve problem solving operations with a number content.

| Factor XI | Spelling | Loading |
|---|---|---|
| 14 | Spelling | .51 |
| 39 | Completion | .42 |
| 20 | Suffices | .37 |
| 9 | Anagrams | .26 |

This factor appeared with the larger verbal factor when fewer factors were extracted. The four tests involve working with the letters of a word and involve a minimum of meaning. It is interesting to note but difficult to interpret that this factor is separate from both the verbal and fluency factors. It is probably a function of the specific experiences of this population where spelling of English is strongly emphasized in schools.

| Factor XII | | Loading |
|---|---|---|
| 45 | Number Code | .65 |
| 6 | Analogies | .31 |
| 38 | Grade-Point Average | .30 |
| 43 | Manila Story Recall | .25 |

In this population of subjects the Number Code test does not load highly the number factor as it did for Thurstone (1938). However both Number Code and Analogies loaded Thurstone's unnamed Factor 11. It would appear that this is a reasoning factor of some sort. There may be an element of understanding the task involved since many subjects had trouble with the instructions for the Number Code test.

| Factor XIII | Word Fluency | Loading |
|---|---|---|
| 4 | Suffixes in Tagalog | .61 |
| 3 | Thing categories in Tagalog | .39 |
| 20 | Suffixes | .34 |

This is a rather poorly defined word fluency factor. The First and Last Letters test in English appears in Factor XIV with a number of memory tests, while the parallel test in Tagalog shows its highest loading on Factor XVII, an unidentified factor with a good many fluency tests.

| Factor XIV | | Loading |
|---|---|---|
| 48 | First and Last Letters | .59 |
| 49 | Number-Number Memory | .48 |
| 40 | First Names | .27 |
| 41 | Word-Number Memory | .25 |

This factor has one fluency and three memory tests. It would appear to involve an ability to recall according to certain rules. For reasons that are not at all clear the First and Last Letters test, ordinarily thought a measure of word fluency, does not load fluency factors for this population. It may be an expression of their memorizing approach to the learning of English.

| Factor XV | Facility with Digits | Loading |
|---|---|---|
| 26 | Digit Symbol | .69 |
| 21 | Writing Digits | .47 |
| 22 | Mathematics Aptitude Test | .26 |

This factor is restricted to material involving digits and appears to represent some facility or confidence in handling them. It may be a product of the methods used in teaching mathematics.

| Factor XVI | Deductive Reasoning | Loading |
|---|---|---|
| 10 | Number Series II | .48 |
| 31 | Verbal Classification | .42 |
| 50 | Paragraph Recall | .41 |
| 27 | Reasoning | .29 |
| 8 | Disarranged Sentences | .28 |

These tests, insofar as they have a common element, appear to involve a deductive reasoning process. There is an element in each of these tests of evaluating the plausibility of alternative solutions or getting a sense of the abstraction involved in the problem.

| Factor XVII | | Loading |
|---|---|---|
| 11 | First and last letters in Tagalog | .58 |
| 2 | Topics in Tagalog | .31 |
| 18 | Thing Categories | .29 |
| 13 | High Numbers | .27 |
| 20 | Suffixes | .25 |

As was the case in the First and Last Letters test in English, the parallel test in Tagalog does not load highly the Word Fluency factor. Four of the five tests on this factor have been found to measure either Word or Ideational Fluency. It would appear that these two fluency factors may not be so sharply differentiated in this population.

| Factor XVIII | Perceptual Speed | Loading |
|---|---|---|
| 29 | First Digit Cancellation | .60 |
| 32 | Identical Forms | .41 |

This factor has been found in many studies.

| Factor XIX | | Loading |
|---|---|---|
| 44 | Flags | .53 |
| 20 | Suffixes | .26 |
| 27 | Reasoning | .26 |

With the elimination of the Cubes test, there was only one test in the battery designed to measure spatial ability. For this reason, a good space factor cannot be expected.

| Factor XX | Span Memory | Loading |
|---|---|---|
| 42 | Digit Span | .59 |
| 40 | First Names | .37 |

This factor is concerned with memory span. Notice that Paragraph Recall or meaningful memory does not load this factor. Other rote memory tests show loadings which approach .25.

| Factor XXI | | Loading |
|---|---|---|
| 36 | Figure Recognition | .61 |
| 50 | Paragraph Recall | .35 |

Figure Recognition is the only visual recall test in the battery. A visual memory factor might have appeared if more tests which required the recall of visually presented material had been included.

| Factor XXII | | Loading |
|---|---|---|
| 22 | Mathematics Aptitude Test | .58 |
| 2 | Topics in Tagalog | .29 |

As it stands, no interpretation of this factor is offered. Interpretation might have been possible if more general reasoning tests had been included. It should be noted that the Mathematics Aptitude Test does not load numerical or other reasoning factors with this group.

## DISCUSSION

There is rather clear evidence that certain factors have emerged in this population which are clearly replications of earlier factors. These include numerical, verbal, rote memory, visualization, motor speed, perceptual speed, and ideational fluency. The cross-cultural status of reasoning factors is obscure. Our Factors IX, X, XII, and XVI are probably reasoning factors but they are not readily matched with results on American subjects. In order to clarify the differences in reasoning, if such exist, it would be necessary to include tests in Tagalog as well as English. The interesting point is that reasoning processes do show at least four factors when tests are given in English to these subjects who for the most part do not habitually think in English.

The emergence of separate vocabulary factors in English and Tagalog parallels the findings of Vandenberg (1959). This is in contrast to the ideational fluency and anagrams factors which include tests in both languages. In view of this central role of the verbal factor in academic progress this finding may have considerable implications for educational exchanges.

Factors XI and XV appear to be products of the Philippine educational system. Spelling of English words is a more specific skill which can be separated from verbal facility. No spelling test was offered in Tagalog since Tagalog is written phonetically in western script. Similarly, facility with digits is differentiated from numerical ability. This is possibly a product of the method by which mathematics was taught to these subjects.

Memory tests were included because it has been said that Philippine education emphasizes memory at the expense of understanding. Four factors emerged: III, XIV, XX, and XXI, which are in whole or part memory factors. Rote memory appears clearly as Factor III, but meaningful memory tests appear on verbal, fluency, and reasoning factors and do not themselves define a factor. Factors XX and XXI are not very convincing instances of span or visual memory factors. Since the Grade-Point Average does not load any of the memory factors, there is little evidence in these results of the prevalence of memory activities said by some to so characterize Philippine education and oriental education generally. It may be that the lower facility in English of these subjects inhibits their spontaneity in that language and so gives rise to the frequent observation that these students rely heavily on memorizing. Parallel tests in Tagalog would possibly clarify some of these issues.

The Word Fluency tests in English and Tagalog do not appear on the same factor. There is much less evidence for a common word fluency factor across languages than for the ideational fluency factor. The First and Last Letters test in English may be more a measure of memory for this group. The same test in Tagalog appears with the heaviest loading on Factor XVII, a factor which is difficult to interpret. We are led to conclude that word fluency as measured by the tests in this battery is a very tenuously defined factor.

From these results we may generalize that the most frequently identified ability factors can be identified in this population. There are, however, verbal factors specific to the language in which the test is given. Other factors emerge which are possibly the product of instructional methods. These factors, obtained on tests given in English, tend to be more specific reflecting the fact that English is not the native language of the subjects and, with lower facility, less generalization of ability has taken place. Finally, all subjects were women. There is little systematic evidence whether the sex of subjects makes a difference in such a study as this. However, the possibility remains that results with male subjects or both sexes would not be the same.

In the same way that cross-culture studies of personality have shed much light on the modifiability of personality, and have offered opportunities to observe different effects on development of different environmental influences, cross-culture studies of the structure of abilities should clarify the role of environmental factors in the structure and development of abilities. Bilingual populations afford an unusual opportunity for this kind of investigation since they can be given equivalent forms in two languages. Reasoning and memory activities represent especially fruitful areas for further investigation. Much remains to be done, however, to determine how equivalent tests are in dissimilar languages and how much a society influences the subject's definition of the testing situation itself.

## REFERENCES

BAKER, H. J. *Detroit Advanced Intelligence Test.* Bloomington, Ill.: Public School Publishing, 1925.

BIESHEUVEL, S. Psychological tests and their application to non-European peoples. In G. B. Jeffrey (Ed.), *The year book of education.* London: Evans, 1949. Pp. 87–117.

BIESHEUVEL, S. The study of African ability: I. Intellectual potential of Africans. *Afr. Stud.*, 1952, **11**, 45–58. (a)

BIESHEUVEL, S. The study of African ability: II. A survey of some research problems. *Afr. Stud.*, 1952, **11**, 105–117. (b)

CARREON, M. L. *Philippine studies in mental measurement.* Yonkers-on-Hudson, New York: World Book, 1926.

CLEETON, G. U. *Carnegie Mental Ability Tests: Form A.* New York: Houghton Mifflin, 1932.

DUNLAP, J. W. Race differences in the organization of numerical and verbal abilities. *Arch. Psychol., N. Y.*, 1931. No. 124.

FRENCH, J. W. *Kit of selected tests for reference aptitude and achievement factors.* Princeton: Educational Testing Service, 1954.

JAHODA, G. Assessment of abstract behavior in a non-Western culture. *J. abnorm. soc. Psychol.*, 1956, **53**, 237–243.

KELLEY, H. P. *A factor analysis of memory ability.* Princeton: Princeton University, 1954.

KLINEBERG, O. *Race differences.* New York: Harper, 1935.

MANN, C. W. A test of general ability in Fiji. *J. genet. Psychol.*, 1939, **54**, 435–454.

NADEL, S. F. The application of intelligence tests in the anthropological field. In F. C. Bartlett (Ed.), *The study of society, methods and problems.* New York: Macmillan, 1939. Pp. 184–198.

PORTEUS, S. D., & BABCOCK, M. E. *Temperament and race.* Boston: Gorham, 1926.

SAUNDERS, D. R. *The contribution of communality estimation to the achievement of factorial invariance.* (Res. Bull. No. 60–5) Princeton: Educational Testing Service, 1960.

SAUNDERS, D. R. Integrating the implementation of quartimax, varimax, oblimax and related rotational procedures. *Psychol. Rep.*, 1962, **10**, 241–242.

SCOTT, G. C. Measuring Sudanese intelligence. *Brit. J. educ. Psychol.*, 1950, **20**, 43–54.

THURSTONE, L. L. Primary mental abilities. *Psychometr. Monogr.*, 1938, No. 1.

THURSTONE, L. L., & THURSTONE, T. F. Factorial studies of intelligence. *Psychometr. Monogr.*, 1941, No. 2.

VANDENBERG, S. G. The primary mental abilities of Chinese students. *Ann. N. Y. Acad. Sci.*, 1959, **79**, 257–304.

VERNON, P. E. Recent investigations of intelligence and its measurement. *Eugen. Rev.*, 1951, **43**, 125–137.

VERNON, P. E. Use of intelligence tests in population studies. *Eugen. Quart.*, 1954, **1**, 221–224.

WHORF, B. L. *Language, thought and reality.* New York: Wiley, 1956.

(Received October 9, 1962)

*Journal of Educational Psychology*
1963, Vol. 54, No. 2, 104–109

# PARENTAL ATTITUDES AND COLLEGE SUCCESS

JOHN E. TEAHAN[1]
*University of Wisconsin, Milwaukee*

Though highly successful academically in high school, ½ of the student Ss of this study (46 males and 44 females) did poorly during their 1st year of college. Although no differences were found in the child rearing attitudes of the students themselves, compared to their daughters the mothers of low achieving girls had stronger attitudes relating to domination and the use of discipline ($p < .01$). In the case of low male achievers, the disparity was primarily between father and son, fathers being more possessive and dominating ($p < .05$). It is suggested that insufficient development in self-sufficiency and independence handicapped the underachievers.

Shaw (1961) in discussing the apparent conflicting results of research in academic underachievement makes the point that too often the conclusions based upon studies of low achievers do not take into consideration, among other things, the sex of the students, the academic level at which they are operating, the methods by which underachievement was determined (i.e., whether by academic grades, achievement test results, or both), and the chronicity of the underachievement. Gough (1952, 1953), on the basis of his studies of successful high school, college, and graduate students, has also concluded that different personality characteristics might be more effective at different stages of the educational process. Thus the conclusions which can be drawn from any study of underachievement must be qualified in terms of many varied factors.

In the present study an investigation was made into the child rearing attitudes of college high and low achievers and their parents. None of these students were chronic underachievers in that all had done extremely well in high school and all had demonstrated high levels of achievement on standardized tests. In many respects these students fit one of Shaw's (1961) definitions of "hidden underachievers" for, in spite of the fact that on the basis of past history they seemed to be performing far above the level of other students, they suddenly fell down in their performance when placed into a new setting (college) where demands differed or competition was greater.

## METHOD

### *Subjects*

These consisted of two groups of college freshmen (46 females and 44 males), their fathers, and their mothers. All students had a history of academic excellence in high school and had graduated in the upper 20% of their classes and all demonstrated high scores on the College Qualification Test (CQT) upon entering the university. Half of these students, 23 females (with a mean CQT of 149.4) and 22 males (with a mean CQT of 162.6), had received at least a 3.0 grade-point average by the end of their freshman year and these were designated as high achievers. The other half of the students, 23 females (with a mean CQT of 147.7) and 22 males (with a mean CQT of 161.1), who were matched to the first groups both on the basis of CQT scores and father's occupational level, had received a less than 2.0 grade-point average and they were designated as low achievers.

---

[1] This study was supported by the Wisconsin Alumni Research Foundation. The investigator also wishes to express his appreciation to those members of the University of Wisconsin, Milwaukee, Student Counseling Center who aided in this research.

## *Materials*

The Parental Attitude Scale (Table 1) was constructed of 45 items taken from the Parental Attitude Survey devised by Shoben (1949). These items were arranged into three subscales: the Possessive (P) scale, the Dominating (D) scale, and the Ignoring (I) scale. A parent scoring high on P scale would agree with an item such as, "In the long run it is better, after all, for a child to be kept fairly close to his mother's apron strings"; a dominating parent would agree with an item like, "It is sometimes necessary for a parent to break the child's will"; while a sample ignoring item is, "Children should not annoy their parents with their unimportant problems." The student, his mother and his father were asked to rate each item on a five-point scale from "strongly agree" to "strongly disagree." Statements in each scale were worded in such a way that in order for a person to receive a high

TABLE 1

PARENTAL ATTITUDE SCALE[a]

| Code[b] | Item | Scores[b] | | | | |
|---|---|---|---|---|---|---|
| 31. (P) | A child should feel a deep sense of obligation always to act in accord with the wishes of his parents. | 4<br>MA | 1<br>SD | 3<br>UN | 5<br>SA | 2<br>MD |
| 32. (P) | The best child is the one who shows lots of affection for his parents. | 5<br>SA | 3<br>UN | 2<br>MD | 4<br>MA | 1<br>SD |
| 33. (D) | Children need some of the natural meanness taken out of them. | 3<br>UN | 1<br>SD | 5<br>SA | 2<br>MD | 4<br>MA |
| 34. (I) | Children who fight with their brothers and sisters are generally a source of great irritation and annoyance to their parents. | 2<br>MD | 5<br>SA | 1<br>SD | 3<br>UN | 4<br>MA |
| 35. (I) | Parents should receive more blame for the misbehavior of their children. | 5<br>SD | 4<br>MD | 3<br>UN | 1<br>SA | 2<br>MA |
| 36. (P) | It is often much better for a child to play at other children's homes than to play at his own home. | 1<br>SA | 5<br>SD | 3<br>UN | 2<br>MA | 4<br>MD |
| 37. (D) | No child should ever set his will against that of his parents. | 3<br>UN | 5<br>SA | 2<br>MD | 4<br>MA | 1<br>SD |
| 38. (I) | Children today expect entirely too much from their parents. | 5<br>SA | 3<br>UN | 1<br>SD | 2<br>MD | 4<br>MA |
| 39. (P) | A child should love his parents above everyone else. | 4<br>MA | 5<br>SA | 1<br>SD | 2<br>MD | 3<br>UN |
| 40. (D) | Mild discipline is best. | 5<br>SD | 4<br>MD | 3<br>UN | 2<br>MA | 1<br>SA |
| 41. (D) | Children should never have to be afraid of their parents. | 3<br>UN | 5<br>SD | 2<br>MA | 1<br>SA | 4<br>MD |
| 42. (I) | Children should learn early what they are free to do and not free to do so that they do not always bother their parents. | 2<br>MD | 3<br>UN | 5<br>SA | 1<br>SD | 4<br>MA |
| 43. (P) | Children should be expected to keep some secrets from their parents. | 2<br>MA | 5<br>SD | 1<br>SA | 3<br>UN | 4<br>MD |
| 44. (D) | Parents cannot expect to have complete control over their children. | 5<br>SD | 1<br>SA | 4<br>MD | 2<br>MA | 3<br>UN |
| 45. (I) | Children too often intrude upon their parents' privacy. | 5<br>SA | 3<br>UN | 1<br>SD | 2<br>MD | 4<br>MA |

[a] For the first 30 items of the scale see Drews and Teahan (1957).

[b] The scores and code letters were not included in the test administration to the subjects.

TABLE 2

POSSESSIVE, DOMINATING, AND IGNORING SUBSCALE COMPARISONS OF FEMALE STUDENTS AND PARENTS

| Subscale | Mean | | | | | |
|---|---|---|---|---|---|---|
| | S | Mo | Fa | s | mo | fa |
| P | 37.4 | 37.8 | 42.4 | 37.8 | 38.1 | 41.9 |
| D | 40.7 | 41.3 | 46.5 | 41.9 | 48.4 | 49.5 |
| I | 36.1 | 36.1 | 36.2 | 38.8 | 40.1 | 41.5 |

| Subscale | *t* ratio | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | S-Mo | S-Fa | Mo-Fa | s-mo | s-fa | mo-fa | S-s | Mo-mo | Fa-fa |
| P | .20 | 2.52* | 1.83 | 1.69 | 2.08* | 1.90 | .25 | .14 | .21 |
| D | .23 | 2.34* | 1.84 | 3.20** | 3.88** | .51 | .59 | 2.70** | 1.23 |
| I | 0. | .69 | .63 | .79 | 1.50 | .72 | 1.70 | 2.21* | 2.99** |

Note.—Abbreviations are: S, high achieving student; s, low achieving student; Mo, mother of high achiever; mo, mother of low achiever; Fa, father of high achiever; and fa, father of low achiever.
* $p \leq .05$.
** $p \leq .01$.

score he would have to agree with half of the items and disagree with the other half. This was to avoid against response set as a determinant in creating a particularly high or low score.

## RESULTS

No differences were found in the child rearing attitudes of high and low achieving female students (see Table 2). However there was a statistically significant difference between the attitudes of mothers and their daughters in the low achieving group. Mothers of low achievers were more dominating than their daughters while no such disparity was found among high achievers. The fathers of both female groups were similar in the sense that they were significantly higher than their daughters on the D and P scales. However, there were qualitative differences between the fathers which were reflected in an item analysis of the scales and these will be reported in the Discussion section. In addition the fathers and mothers of high females were significantly lower on I scale than their low achiever counterparts.

While among the females it was mother's disparate attitudes which seemed to separate high and low achievers, among the males it was primarily father's attitudes which appeared to set these two groups apart (see Table 3). Fathers of low achievers were significantly higher than their sons on the P and I scales while no such differences appeared when the high achievers and their fathers were compared. Fathers of high achievers were also significantly lower on P scale than fathers of low achievers. Only in the case of high achievers did a difference appear between sons and mothers, with the latter having a significantly lower score than their offspring on I scale.

## DISCUSSION

As mentioned previously Gough (1952, 1953) concluded on the basis of his studies that in spite of a certain overlap, different personality characteristics might be more effective at different stages of the educational process. He found, for example, that the successful student in high school was very accept-

TABLE 3

POSSESSIVE, DOMINATING, AND IGNORING SUBSCALE COMPARISONS OF MALE STUDENTS AND PARENTS

| Subscale | Mean | | | | | |
|---|---|---|---|---|---|---|
| | S | Mo | Fa | s | mo | fa |
| P | 40.5 | 40.4 | 38.4 | 38.8 | 40.2 | 43.8 |
| D | 43.7 | 45.9 | 45.2 | 43.7 | 47.4 | 50.5 |
| I | 41.5 | 38.0 | 41.2 | 41.3 | 38.7 | 39.9 |

| Subscale | *t* ratio | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | S-Mo | S-Fa | Mo-Fa | s-mo | s-fa | mo-fa | S-s | Mo-mo | Fa-fa |
| P | .50 | 1.21 | .89 | .59 | 2.05* | 1.52 | .98 | .09 | 2.45* |
| D | .83 | .65 | .28 | 1.29 | 2.27* | 1.02 | 0. | .52 | 1.94 |
| I | 2.06* | .15 | 1.47 | 1.34 | .69 | .60 | .12 | .37 | .58 |

Note.—Abbreviations are: S, high achieving student; s, low achieving student; Mo, mother of high achiever; mo, mother of low achiever; Fa, father of high achiever; and fa, father of low achiever.
* $p \leq .05$.

ing of convention, orderly, and docile. However in college a different constellation of traits seemed to characterize the high achiever with more emphasis on self-sufficiency and independence of judgment. This latter finding agrees with other studies of college students (Lum, 1960; Stern, Stein, & Bloom 1956). Drews and Teahan (1957) in studying the child rearing attitudes of mothers of junior high school students found that the mothers of *high achievers* were more dominating and were stricter disciplinarians than the mothers of low achievers. This finding agrees with Gough's conception of the docile, conforming, successful high school student. Hence the disparity now seen in the present study between low achievers and their mothers (especially in the case of females) and fathers (especially in the case of males) could be interpreted as meaning that these students have not been allowed to develop the self-sufficiency and independence of thought necessary for *college* success. It might also imply that although some level of parental demands and perhaps even domination is necessary, too much can interfere with academic success beyond the secondary school level.

It is interesting to note that one of the most frequent complaints made by college students who have failed to live up to their previous level of academic success in high school regards the fact that no one chases after them in college demanding that assignments be done or deadlines met. One gains the impression that they have learned to lean too heavily upon external forces and when these are not forthcoming they are unable to provide themselves with the necessary controls.

The fact that there was a large discrepancy between the female students and their fathers on P and D scales would seem to indicate that all fathers are somewhat possessive and dominating in their attitudes toward their daughters. However, there appeared to be some qualitative differences when individual items were studied. Both parents of low achievers seemed to demand rather unquestioning obedience from their children and were more punitive in their approach to them. Thus the mothers and fathers of low females were significantly higher than their daughters on such items as, "A child should always believe what his parents tell him" ($p < .01$ for mothers, $p < .04$ for fathers) and "Children need some of the natural meanness taken out of them" ($p < .01$ for mothers, $p < .05$ for fathers). Although the fathers of high achievers felt that their children should not choose friends without restrictions ($p < .05$) and would limit their children to making only minor decisions for themselves ($p < .01$), neither these fathers nor their wives agreed with the more punitive attitudes seen in the parents of lows. The latter felt that harsh discipline was necessary more often than did the parents of highs and this result was also significant at the .02 level of confidence in the case of both mothers and fathers.

The fact that in the case of low achieving females both mother and father are so high on D scale would also possibly indicate that the dominating attitude of the high achiever's father is tempered in the home by more permissiveness on the mother's part with, perhaps, a better relationship between mother and daughter because of their similar attitudes. The low achieving female, on the other hand, faces domineering attitudes from both parents. In addition the fathers of high females were significantly lower on I scale than the fathers of lows who, unlike their daughters, felt that, "A child should be seen and not heard" ($p < .05$) and "Children should learn early what they are free to do and not free to do so that they do not always bother their parents" ($p < .01$). In contrast to this the high fathers agree significantly more often than the fathers of low achievers with the proposition that, "The most important consideration in planning the activities of the home should be the needs and interests of the child" ($p < .05$).

Another interesting feature which emerged from the study was that while in the case of the females it was the disparity between mother-daughter attitudes towards domination which seemed related to the daughter's academic achievement, in the case of the males it was father's attitudes which seemed to play the most significant role. The one exception to this was that mothers of high males seemed to be even more child oriented than their sons. Low fathers, however, were significantly higher than their sons in feeling that children should make only minor decisions for themselves ($p < .02$), that they should always believe what their parents tell them ($p < .04$) and that they should always be under the complete control of their parents ($p < .03$). There was no such disparity between the high fathers and their sons. Low fathers believe even more strongly than their wives that it is wicked for a child to disobey it's parents ($p < .01$) and in contrast to their sons and unlike the high fathers they also feel that parents are generally too busy to answer all of a child's questions ($p < .04$) and that children should not interrupt adult's conversation ($p < .04$). They are also more possessive than their sons in wanting to play a role in the selection of playmates ($p < .04$) and they feel more strongly than either their sons or the high fathers that their children should play at home rather than at the homes of other children ($p < .03$ in both cases). Thus the picture that seems to emerge in the case of the low male achiever is a clash between a domineering, punitive, and overprotective father and his son. One might also, of course, speculate as to whether the disparity seen between the attitudes of low achieving girls and their mothers and between low achieving males and their fathers has any bearing upon a possible lack of identification with these parental figures.

One last comment should be made on the fact that no differences appeared between high and low achieving students in any of the three subscales. Apparently, when sex is held constant, they agree quite well among themselves in terms of expressed attitudes towards child rearing. In some respects this is surprising, for one might expect that those students who have clashed with their parents on issues regarding child rearing might also react with demands for greater freedom or less domination in a parent-child relationship and that this would be reflected in their test results. Perhaps peer influence or the more intellectualized standards learned in college courses have a stronger effect on such attitudes, especially when it pertains to "children in the abstract." Whether these attitudes will hold true when they themselves become parents is, of course, another matter.

## REFERENCES

DREWS, ELIZABETH M., & TEAHAN, J. E. Parental attitudes and academic achievement. *J. clin. Psychol.*, 1957, **13**, 328–332.

GOUGH, H. G. Some theoretical problems in the construction of practical assessment devices for the early identification of high level talent. Berkeley: University of California, Department of Psychology and Institute for Personality Assessment and Research, 1952. (Mimeo)

GOUGH, H. G. What determines the academic achievement of high school students. *J. educ. Res.*, 1953, **46**, 321–331.

LUM, MABEL K. M. A comparison of under- and overachieving female college students. *J. educ. Psychol.*, 1960, **51**, 109–114.

SHAW, M. C. Definition and identification of academic underachievers. In L. M. Miller (Ed.), *Guidance for the underachiever with superior ability*. (Bull. No. 25) Washington, D. C.: United States Government Printing Office, 1961. Pp. 15–30.

SHOBEN, E. J., JR. The assessment of parental attitudes in relation to child adjustment. *Genet. Psychol. Monogr.*, 1949, **39**, 103–148.

STERN, G. C., STEIN, M. I., & BLOOM, B. S. *Methods in personality assessment*. Chicago: Free Press, 1956. Pp. 161–184.

(Received October 12, 1962)

*Journal of Educational Psychology*
1963, Vol. 54, No. 2, 110–117

# CORRELATION OF TEACHER PERSONALITY VARIABLES AND STUDENT RATINGS[1]

ROBERT L. ISAACSON, WILBERT J. McKEACHIE, AND JOHN E. MILHOLLAND
*University of Michigan*

A peer group nomination procedure, a descriptive adjective inventory, and the IPAT 16 PF Questionnaire were administered to 2 groups of teaching fellows in the introductory psychology course at the University of Michigan. These personality scores were related to the teaching fellows' teaching effectiveness ratings made by their students and to the factor scores obtained from factor analyses of the separate items of the student rating instrument. The teacher personality variable most consistently correlated with good ratings by students was the peer group evaluation of the teaching fellows' "general cultural attainment." The relations of the personality scales to the other dimensions of student ratings are also presented.

Stable correlates of teaching success have been hard to find. Studies of teacher personality characteristics related to successful teaching have often yielded conclusions which needed stringent qualifications. For example, Della Piana and Gage (1955) found that students' evaluations of teachers were related to the teachers' scores on the Minnesota Teacher Attitude Inventory (MTAI) *if* the classes were oriented generally toward the affective merits of the teachers. The student ratings of classes in which the predominant orientation was not in this direction did not correlate with the teachers' MTAI scores. These findings suggest an explanation for the lack of correlation between the MTAI teacher scores and student evaluations when the students involved were all enrolled in a university high school and presumably more oriented toward cognitive rather than affective teacher qualities (Gage & Suci, 1951).

Gowan (1955) stated that persons making high scores on the *K* scale of the Minnesota Multiphasic Personality Inventory (MMPI) show a tendency to enter into empathic relationships and to make good teachers and counselors. In addition, teachers identified as well-adjusted were thought to be the better teachers. Symonds and Dudek (1956), evaluating personality with the use of the Rorschach, reported that superior personal organization, good adjustment and reasoning, capacity to relate to others, and less aggressive responses correlated with successful teaching. In his recent book Ryans (1960) contrasted traits characteristic of teachers high and low on three scales of desirable teaching characteristics. Among the traits distinguishing the good teacher were emotional stability, friendliness, cooperativeness, agreeableness, restraint, and objectivity.

Practically the only agreement among researchers on teacher characteristics is that all of the traits of effective teachers are valued in our culture, although we should pause to note the suggestion that neurotic individuals *can* be effective, given appropriate circumstances of work (see the discussion of Symonds, 1950, pp. 692–693).

It may be that only certain traits from clusters of ones with high cultural evaluation will predict success in college instruction. At the time the study reported in this paper was undertaken, a technique of assessment of five culturally valued traits was being investigated by Norman (1961). This technique uses peer group nominations on 20 individual bipolar scales which are combined to yield scores on five personality factors. It stemmed from the work of Cattell (1947) and was further developed by Tupes (1957) and Tupes and Christal (1958, 1961). The factors defined by the peer nomination procedure are labeled:

---

[1] This research was supported in part by Office of Education Research Contract O. E. No. SAE-8451, to W. J. McKeachie, J. E. Milholland, and R. L. Isaacson.

1. Surgency (assertive, frank, energetic, talkative, etc.)
2. Agreeableness (cooperative, attentive, mild-mannered, not jealous, etc.)
3. Dependability (responsible, conscientious, orderly, etc.)
4. Emotional Stability (calm, poised, in control of emotions, etc.)
5. Culture (artistic, polished, imaginative, effectively intelligent, etc.)

Because all five factors describe traits relevant to teaching we believed that the peer nominations might be useful in identifying specific personality traits correlated with effective college teaching.

## PROCEDURE

### *Teacher Sample*

To obtain a sample of college-level instructors who were relatively homogeneous in age, general abilities, teaching experience, and instructional media, we secured the participation of the teaching fellows in the introductory psychology course at the University of Michigan. In 2 successive years two different groups of young men and women who had not taught at the University of Michigan before became available.

In their initial instructional experience the teaching fellows led discussion sections of a large lecture course. In the second semester each was entirely responsible for his own class and both lectured and led discussion sections. We could therefore study the personality variables related to success under two types of course organization, although the second semester experience was necessarily confounded with experience derived from the first term.

All of the teaching fellows in the introductory psychology course during the 1959–60 and the 1960–61 academic years cooperated in this study. The majority was advanced graduate students in psychology. All had successfully completed their first year of graduate study and had passed the departmental Master's Degree Qualifying Examination or its equivalent. In each of the two teaching fellow groups one of the instructor-subjects was a woman. No teaching fellow was an instructor-subject both years. To avoid contamination of the 1960–61 group by knowledge of previous results, no information about 1959–60 results was circulated. There were 17 teaching fellows in the fall of 1960–61 and 16 in the spring of 1960–61.

### *Student Subjects*

All students enrolled in the introductory psychology courses were asked to fill out instructor rating forms. Since it was contrary to department policy to require students to do the ratings if they did not wish to, complete returns could not be secured. Furthermore, not all of the completed forms could be used in our analyses because some students responded in an ambiguous fashion or failed to answer certain items. The ratings of the teaching fellows represent the mean of between 35 and 40 student evaluations collected from all of the sections taught by a given teaching fellow. Usable evaluations forms varied from 58% to 84% of the total enrollments in the semesters reported.

### *Measures*

The peer group nomination form developed by Norman (1961) was administered to the teaching fellows in a teaching seminar they attended. Each teacher was asked to place a third of his fellow teachers at each pole of 20 items defined by "polar description" and covering the five traits referred to earlier. The remaining third on each item were then assigned to a middle group. The score for a teaching fellow on a trait was obtained by subtracting the number of negative pole nominations he received from the number of positive pole nominations he received on the four items defining a given trait.

A self-descriptive adjective checklist, also developed by Norman (1961), was also filled out by the teaching fellows during a meeting of the seminar. This inventory consists of 200 pairs of adjectives equated for social desirability. The respondent is required to indicate which of a pair of adjectives more closely describes him. The five scores obtained from this form correspond to the five traits covered by the peer group nomination form.

In addition each teaching fellow completed Forms A and B of the Institute for Personality and Ability Testing Sixteen Personality Factor (16 PF) Questionnaire (Cattell, Saunders, & Stice, 1957).

The students in the four semesters studied responded to two different evaluation forms. The form used in the year 1959–60 and the fall semester 1960–61 was a short form of 12 questions. The form used in the spring term 1961 consisted of 43 items which were selected to obtain a greater number of teacher-evaluation factors than were obtained on the shorter form. Factor analysis of the short form produced a large general factor. The item loading highest on this factor was:

How would you rate your instructor in general (all-around) teaching ability?

1. An outstanding and stimulating teacher
2. A very good instructor
3. A good instructor
4. An adequate but not stimulating instructor
5. A poor and inadequate instructor

The longer form used in the spring of 1961 contained several items from the shorter form. The item described above was one of these and it loaded highest on the most prominent factor obtained from factor analyzing this new form. Subsequent study of student evaluations with this longer instrument has revealed five consistent factors. In the present paper personality scores of the teaching fellows will be related to the students' evaluation of them on the "overall effectiveness" items, given above, for all semesters and related to the teaching fellow factor scores for the five factors.

## RESULTS

The students' responses to the overall teaching effectiveness item were scored 1–5, from most to least desirable, and the means for the various teaching fellows correlated with their scores on each of the 26 personality measures. The correlations are given in Table 1.

The only consistently high relationship is that of Peer Group Nomination Factor 5, Culture, with the students' evaluation. The relation appears in both groups and in all four semesters studies, and leads us to believe that this variable is important. This finding has been replicated at the Ohio State University,[2] where the correlation between peer ratings of culture and student ratings of effectiveness was .48.

Factor scores were obtained for each teaching fellow on the five factors resulting from the analysis of the student rating instrument used in the spring 1961. These factors and the items which loaded highly on them are presented in Table 2. The labels for the factors are somewhat arbitrary, but are intended to reflect what the general character of the item cluster is.

The significant correlations of the personality scores of the teaching fellows with the factor scores for each of the five factors are presented in Table 3.

## DISCUSSION

Since the Peer Group Nomination Scale provides the most stable predictor of the students' rating of a teacher's overall effectiveness, it behooves us to look in detail at its "culture" factor and the procedures used to obtain it. The four subscales over which scores are summed to obtain a total Culture score are characterized by the following polar descriptions:

| | |
|---|---|
| Artistically sensitive | Artistically insensitive |
| Intellectual | Unreflective, narrow |
| Polished, refined | Crude, boorish |
| Imaginative | Simple, direct |

Teaching fellows rated by their peers as artistically sensitive, intellectual, and so on, tend to be the ones rated as effective teachers by their students. How should this be interpreted?

---

[2] The authors express their appreciation to D. Wickens who administered the Peer Nomination Form and student rating form at Ohio State University.

First of all we should note that because this is a peer nomination procedure we have no information as to whether the teachers with superior student evaluations are *really* more artistically sensitive, intellectual, etc. We may conclude only that the other teaching fellows *see* them as more "cultured" than the less favorably rated members of the group. It remains for further research to determine whether this impression of greater cultural attainment is really based on greater cultural skills.

The students, themselves, tend to see

TABLE 1

CORRELATIONS[a] BETWEEN PERSONALITY VARIABLES AND STUDENT RATINGS OF OVERALL TEACHING ABILITY

| Personality variable | Semester | | | |
|---|---|---|---|---|
| | 1959–60 | | 1960–61 | |
| | Fall | Spring | Fall | Spring |
| Peer Group Nominations | | | | |
| 1. Surgency | .40 | .41 | .58* | .66** |
| 2. Agreeableness | .61* | .66** | .52* | −.01 |
| 3. Dependability | .41 | .36 | −.06 | −.15 |
| 4. Emotional Stability | .35 | .47 | .58* | .75 |
| 5. Culture | .60* | .65** | .67** | .48 |
| DAI | | | | |
| 1. Surgency | .12 | .23 | .46 | .51* |
| 2. Agreeableness | .07 | −.03 | .01 | −.03 |
| 3. Dependability | −.33 | −.25 | −.35 | −.48 |
| 4. Emotional Stability | .01 | .07 | .53* | −.42 |
| 5. Culture | .27 | .13 | .25 | .25 |
| Cattell factors | | | | |
| 1. Warmth | .19 | .36 | .36 | .13 |
| 2. Brightness (g) | −.06 | −.03 | −.04 | −.37 |
| 3. Mature, calm | −.07 | −.05 | −.48* | −.30 |
| 4. Aggressiveness | .42 | .50* | .27 | .45 |
| 5. Enthusiasm | .52* | .56* | .43 | .51* |
| 6. Conscientiousness | −.58 | −.56* | −.05 | −.28 |
| 7. Adventurous | .46 | .46 | .23 | .01 |
| 8. Sensitivity | .25 | .50* | .02 | .18 |
| 9. Jealous, suspecting | .00 | −.03 | −.19 | −.03 |
| 10. Bohemian | .33 | .43 | −.01 | .47 |
| 11. Sophisticated | −.31 | −.34 | −.35 | −.42 |
| 12. Anxious, insecure | .05 | .05 | .08 | .20 |
| 13. Experimenting, critical | .43 | .47 | .14 | .44 |
| 14. Self-sufficient | −.17 | −.08 | −.42 | .02 |
| 15. Controlled, precise | −.23 | −.29 | .03 | −.44 |
| 16. Tense, excitable | −.21 | −.16 | .31 | .24 |
| *N* | 16 | 16 | 17 | 16 |

[a] Positive correlations designate high scores on personality variables, as named, going with good student ratings.

* $p < .05$.

** $p < .01$.

TABLE 2

FIVE FACTORS OBTAINED FROM THE STUDENT EVALUATIONS IN THE SPRING OF 1961 AND THE ITEM WHICH LOADED HIGHLY ON THEM

| Factor and Item | Loading | |
|---|---|---|
| | Men students | Women students |
| I. Skill | | |
| Put materials across in interesting way | .863 | .847 |
| All-around teaching ability | .818 | .859 |
| Stimulated intellectual curiosity | .755 | .812 |
| Explained things clearly and concisely | .782 | .706 |
| Skillful in observing student reactions | .568 | .645 |
| Changed approach to meet new situations | .689 | .674 |
| Tried to increase interest of class in subject | .640 | .487 |
| Class time well spent | .626 | .674 |
| Made clear how each topic fit into course | .561 | .471 |
| Overall value of course | .341 | .589 |
| Explained why he did things | .556 | .442 |
| Anticipated difficulties before they arose | .572 | .550 |
| Improved ability to understand deviant individuals | .534 | .423 |
| Improved ability to deal with personal problems | .485 | .528 |
| Was aware when students failed to keep up | .281 | .348 |
| Explained reasons for his criticism | .560 | .463 |
| High degree of cultural attainment | .574 | .568 |
| II. Overload | | |
| Assigned a great amount of reading | .704 | .666 |
| Assigned difficult reading | .679 | .699 |
| Asked for more than students could get done | .537 | .675 |
| III. Structure | | |
| Planned class activities in detail | .585 | .727 |
| Had everything going on schedule | .613 | .698 |
| Followed an outline closely | .645 | .680 |
| Decided in detail what and how things should be done | .417 | .598 |
| IV. Student Rapport | | |
| Invited criticisms of his acts | .463 | .250 |
| Told students when they had done a good job | .428 | .366 |
| Complimented students in front of others | .424 | .490 |
| Able to get personal help in course | .315 | .199 |
| Improved ability to deal with personal problems | .184 | .322 |
| Criticized poor work | −.006 | .225 |
| Was sensitive to students' desire to ask questions | −.010 | −.016 |
| Anticipated difficulties before they arose | .283 | .171 |
| Permissive and flexible | .075 | .108 |
| V. Group Interaction | | |
| Students argued with each other and instructor without hostility | .383 | .486 |
| Students frequently volunteered own opinions | .302 | .494 |
| Students felt free to express own opinions | .556 | .280 |
| Students were friendly | .384 | .189 |
| Instructor was aware when students failed to keep up | .425 | .012 |

teachers they judge to be effective as possessing high cultural attainment. An item asking the students to rate the extent to which the instructor exhibited a high degree of cultural attainment correlated .48 with the criterion item of overall teaching effectiveness. Items from the student rating form asking for evaluation of the teachers' talkativeness correlated approximately .36 with the criterion overall effectiveness item, and an item asking for evaluation of the instructors' permissiveness and flexibility correlated only .24 with the same criterion item.

The fact that this appearance of general cultural knowledgeability should relate to the student ratings is not without precedent. For example, Cosgrove (1959) found students' opinions of the extent to which instructors were "well-read" and informed about material in related fields to load on a factor best labeled as an index of mastery and organized of course content. Beardslee and O'Dowd (1961) also found the student

TABLE 3

CORRELATIONS BETWEEN PERSONALITY VARIABLES AND FACTOR SCORES OBTAINED FROM STUDENT RATINGS

| Men's student rating factors and personality variable | r | Women's student rating factors and personality variable | r |
|---|---|---|---|
| I. Skill | | I. Skill | |
| PG 1—Surgency | .69 | PG 1—Surgency | .55 |
| PG 4—Emotional Stability | .77 | PG 4—Emotional Stability | .66 |
| PG 5 Culture | .58 | PG 5—Culture | .47 |
| DAI 1—Surgency | .54 | DAI 1—Surgency | .46 |
| DAI 1—Dependability | −.50 | DAI 4—Emotional Stability | −.60 |
| 16 PF (F) Enthusiastic, talkative | .53 | 16 PF (E)—Dominant, aggressive | .49 |
| 16 PF ($Q_1$)—Experimenting, critical, radical | .50 | 16 PF (F)—Enthusiastic, talkative | .57 |
| II. Overload: No correlates | | 16 PF (N)—Sophisticated, polished | −.56 |
| III. Structure | | II. Overload: No correlates | |
| 16 PF ($Q_3$)—Lax, unsure | −.48 | III. Structure | |
| IV. Student Rapport | | 16 PF ($Q_3$)—Lax, unsure | −.50 |
| PG 1—Surgency | .53 | IV. Student Rapport | |
| DAI 3—Dependability | −.50 | PG 1—Surgency | .48 |
| 16 PF ($Q_1$)—Experimenting, critical, radical | .72 | PG 4—Emotional Stability | .54 |
| V. Group Interaction | | PG 5—Culture | .52 |
| PG 1—Surgency | .55 | V. Group Interaction | |
| PG 5—Culture | .49 | PG 1—Surgency | .74 |
| | | PG 4—Emotional Stability | .46 |
| | | PG 5—Culture | .65 |
| | | DAI 4—Emotional Stability | −.53 |
| | | DAI 5—Culture | .49 |
| | | 16 PF (C)—Mature, calm | −.62 |
| | | 16 PF (M)—Bohemian, unconcerned | .50 |
| | | 16 PF ($Q_4$)—Tense, excitable | .50 |

Note.—Only those correlations which exceed the .05 level of confidence are reported. Abbreviations: PG, Peer Group Nominations; DAI, Descriptive Adjective Inventory; 16 PF, Cattell factors.

image of college professors emphasized their all-around intellectual competence and sensitivity to artistic and esthetic experience. College professors who overtly manifest these characteristics were perceived by students as closer to the ideal than those who do not.

There may be some question, however, whether this salience of a general culture factor will generalize to success in teaching in areas other than psychology. The work of Riley, Ryan, and Lifshitz (1950) indicates that there may be some differences among what students think are most important for instructors in the arts, the sciences, and the social sciences.

In our study the patterns of correlation for the other variables with the effectiveness item were inconsistent. In three out of the four semesters, however, the Peer Group Nomination variables Agreeableness, and Emotional Stability, and the 16 PF Enthusiasm score correlated well with the student evaluations. We have no hesitancy in saying that a college teacher's possession of agreeableness, emotional stability, and enthusiasm, as well as high cultural attainment, augurs well for him. The preceding statement seems in fact to be more a validation of our measures than a discovery of the attributes of a successful teacher.

Some of the variables that failed to make the grade excite our interest. Do students care whether a teacher is dependable? Would they just as soon have "cold" teachers as "warm" ones? In another study McKeachie (1961) found warmth, as rated by students and observers, to relate to achievement differently for men and women and for students high and low in need affiliation. If warmth affects some students positively and others negatively, the effect may cancel out in a simple correlational study of the sort reported here. Other variables may interact in similar ways.

The Descriptive Adjective Inventory (DAI) scores yielded only two significant correlations with the overall effectiveness item. The tendency for the Emotional Stability scores in 1960–61 to be negatively correlated may be of some moment. Since the correlations for this score evaluated via Peer Group Nominations are positive it may mean that in this group of teachers the ones who were regarded by their peers as most stable were disposed to regard themselves as unstable, and vice versa. If this sort of lack of correspondence is characteristic it may undermine the utility of the self-report instrument. Norman's (1962) results, however, may be interpreted more optimistically, since he found a correlation of .40 between the Emotional Stability scores in the two instruments.

When personality variables are considered in relation to the student rating factors, the same pattern described above for the students' responses to the overall effectiveness item is found for Factor I for both sexes. There are, however, some different personality correlates of the Skill factor for men and for women (see Table 3).

No significant personality correlates were found for Factor II (Overload). The factor analysis of the student rating form did, however, separate this overload factor, which deals with assignments, from one which describes an authoritarian class structure, Factor III. The Cattell personality dimension ($Q_3$) correlates with this student rating factor in such a way that the person who describes himself as lax and unsure of himself is the one who is compulsive in classroom structure.

Several personality variables correlate with the student rating factor called Student Rapport, but only one, Surgency, appears for students of both sexes. The peer group "surgency" dimension reflects the talkative and enthusiastic individual who seems to be perceived as being able to get along well with the students. It appears as a correlate of the student rating factor Group Interaction also, along with Culture.

We regard the relations we found between the teacher personality variables and the student rating factors as established only tentatively, but in the main they seem reasonable. We have not attempted as yet to argue the desirability of high standing on any of the factors, but we would be inclined to favor Skill, Student Rapport, and Group Interaction. Under these circumstances, teachers high on Surgency, Culture, and Emotional Stability (as judged by their peers, not by themselves) stand out as most likely to be seen in a good light by their students.

## REFERENCES

BEARDSLEE, D. C., & O'DOWD, D. D. Higher education and the occupational world. In N. Sanford (Ed.), *The American college*. New York: Wiley, 1961.

CATTELL, R. B. Confirmation and clarification of primary personality factors. *Psychometrika*, 1947, **12**, 197–220.

CATTELL, R. B., SAUNDERS, D. R., & STICE, G. *Handbook for the Sixteen Personality Factor Questionnaire*. Champaign, Ill.: Institute for Personality and Ability Testing, 1957.

COSGROVE, D. S. Diagnostic rating of teacher performance. *J. educ. Psychol.*, 1959, **50**, 200–204.

DELLA PIANA, G. M., & GAGE, N. L. Pupils' values and the validity of MTAI. *J. educ. Psychol.*, 1955, **46**, 167–178.

GAGE, N. L., & SUCI, G. Social perception and teacher-pupil relationships. *J. educ. Psychol.*, 1951, **42**, 144–152.

GOWAN, J. C. Relation of the "K" scale on the MMPI to teaching personality. *Calif. J. educ. Res.*, 1955, **6**, 208–212.

McKEACHIE, W. J. Motivation, teaching methods, and college learning. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1961*. Lincoln: Univer. Nebraska Press, 1961. Pp. 111–146.

NORMAN, W. T. Development of self-report tests to measure personality factors identified from peer nominations. *USAF ASD tech. Note*, 1961, No. 61-44.

NORMAN, W. T. Valuation of personality tests as measures of trait-rating factors *USAF PRL tech. docum. Rep.*, 1962, No. 62-4.

RILEY, J. W., RYAN, B. F., & LIFSHITZ, M. *The student looks at his teacher*. New Brunswick: Rutgers Univer. Press, 1950.

RYANS, D. G. *Characteristics of teachers: Their description, comparison, and appraisal.* Washington, D. C.: American Council on Education, 1960.

SYMONDS, P. M. Reflections on observations of teachers. *J. educ. Res.*, 1950, 63, 688–696.

SYMONDS, P. M., & DUDEK, S. Use of Rorschach in diagnosis of teacher effectiveness. *J. proj. Tech.*, 1956, **20**, 227–234.

TUPES, E. C. Relationship between behavior trait ratings by peers and later officer performance of USAF Officer Candidate School graduates. *USAF PTRC res. Rep.*, 1957, No. 57–125.

TUPES, E. C., & CHRISTAL, R. E. Stability of personality trait rating factors obtained under diverse conditions. *USAF WADC tech. Note*, 1958, No. 58-61.

TUPES, E. C., & CHRISTAL, R. E. Recurrent personality factors based on trait ratings. *USAF ASD tech. Rep.*, 1961, No. 61-97.

(Received November 9, 1962)

*Journal of Educational Psychology*
1963, Vol. 54, No. 2, 118–122

# ACADEMIC PERFORMANCE OF PUBLIC AND PRIVATE SCHOOL STUDENTS

JOHN A. FINGER
*Brown University*

AND

GEORGE E. SCHLESSER
*Colgate University*

Public school students are reported to obtain higher college grades than private school students. Equating on scholastic aptitude is reported to leave a difference favoring the public school student. The Personal Values Inventory, a test measuring nonintellectual factors in school success, was administered to an entering college class. 216 had attended public school and 97 private school. When grades were regressed for the Scholastic Aptitude Test, Verbal and Mathematical; the Ohio State Psychological Examination; and the Personal Values Inventory, the group means did not differ significantly. Regressing for these factors randomly distributes grades among private and public school students. A second sample showed similar results for the group means. The conclusion is that private school students do less well because as a group they are lower in scholastic aptitude and in motivation.

Many data substantiate the conclusion that public school students receive higher grades in college than do private school students (Lathrop & Kieffer, 1959; McArthur, 1954; Seltzer, 1948; Shuey, 1956, 1958). Various hypotheses have been advanced to account for this difference.

One hypothesis is that public school students do better in college because they are more rigorously selected for admission. This hypothesis can be dealt with in two parts, for selective factors may be based on either scholastic aptitude—that is, intellectual factors—or on motivational, or nonintellectual, factors. Equating groups for differences in aptitude has become a standard practice in educational research. Similar equating of groups for motivational differences has not yet become common because of lack of adequate test instruments.

It will be shown that when public and private school students are equated for both academic aptitude and academic motivation, most, if not all, of the differences between the two groups are accounted for.

The data will be shown to be consistent with the affirmative hypothesis that differences in performance between public and private school students when found are due to selection procedures.

The procedures utilized in this paper have broader implications than the equating of public and private school students. Differences between many other groups may disappear when both intellectual and nonintellectual factors are held constant. Accounting for motivational differences will enable hypotheses to be more definite in scope, for hypotheses can be confined to those concerning factors related to: (*a*) aptitude differences between groups, (*b*) motivational differences between groups, (*c*) neither aptitude nor motivational differences between groups.

## PROCEDURE

A test of academic motivation, the Personal Values Inventory[1] (PVI; Schlesser & Finger,

---

[1] A copy of the Personal Values Inventory has been deposited with the American Documentation Institute. Order Document No. 7463 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. D., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

1961), has been devised at Colgate University. Scores on this test are available on large populations of college students who have served as the norming groups. The Inventory contains items designed to measure student attitudes toward school work and peer group activities. It was constructed by selecting items which correlate with grade-point average (GPA) but which do not correlate with scholastic aptitude. The correlation of the PVI score with college grades for this sample is .40; PVI with the Scholastic Aptitude Test, Verbal (SAT-V) is .09. The correlation of PVI with grades with SAT-V held constant is .39. These correlations are representative of those obtained in a sample testing of entering freshmen in approximately 30 colleges.

Sample items are:

1. My friends often tried to discourage others from working hard.

2. I and most of my friends have seemed to be in rebellion against our parents.

3. At times have you felt rather comfortably lazy?

4. Do you have more difficulty keeping your mind on your work than others seem to have?

For many research purposes where motivational differences should be controlled the high school record which is a correlate of both aptitude and motivation can be used as a substitute for a test of motivation. In the context of the problem here such a procedure is obviously not possible.

The subjects in this study were the 313 Colgate University students, of whom 216 had attended public schools and 97 private schools. They are described in Table 1 which shows the number receiving various GPAs in the first semester over the past 3 years. As will be noted, public school students consistently earn higher grades.

Table 2 shows the test scores and performance levels of public and private school students on PVI; the College Entrance Examination Board, SAT-V and SAT-M; and the Ohio State Psychological Examination (OSPE); and GPA for the student's first semester in college. Included in Table 2 are five *F* ratios. In all five instances the differences between public and private school students are shown to be significant at the 1% level. On the basis of Table 2 it can be concluded that the groups differ on the three separate measures of scholastic aptitude, on academic motivation as measured by PVI, and on GPA.

The reason for the difference in performance is hypothesized to be attributable to differences in scholastic ability and differences in academic

TABLE 1

Public and Private School Students Receiving Various First Semester GPAs at Colgate University

| First semester GPA | Type of school | Number of students by entrance year | | | | |
|---|---|---|---|---|---|---|
| | | 1959 | 1960 | 1961 | Total | Total % |
| 3.2 or higher | Public | 30 | 39 | 25 | 94 | 14.4 |
| | Private | 1 | 5 | 3 | 9 | 3.6 |
| 2.8–3.1 | Public | 38 | 29 | 38 | 105 | 16.1 |
| | Private | 9 | 14 | 13 | 36 | 14.5 |
| 2.4–2.7 | Public | 45 | 54 | 42 | 141 | 21.6 |
| | Private | 18 | 21 | 9 | 48 | 19.3 |
| Total % high | Public | | | | | 52.1 |
| | Private | | | | | 37.5 |
| 1.8–2.1 | Public | 57 | 65 | 51 | 173 | 26.5 |
| | Private | 27 | 31 | 23 | 81 | 32.7 |
| 1.4–1.7 | Public | 35 | 22 | 38 | 95 | 14.6 |
| | Private | 18 | 14 | 22 | 54 | 21.8 |
| Under 1.4 | Public | 6 | 18 | 20 | 44 | 6.7 |
| | Private | 0 | 6 | 14 | 20 | 8.1 |
| Total % low | Public | | | | | 47.9 |
| | Private | | | | | 62.5 |

Note.—Four-point scale.

TABLE 2

TEST SCORES AND PERFORMANCE LEVELS OF PUBLIC AND PRIVATE SCHOOL STUDENTS FOR, FRESHMAN CLASS ENTERING IN 1961

| Type of school | *N* | Mean score | | | | |
|---|---|---|---|---|---|---|
| | | Entrance examinations | | College examinations | | Grade first semester |
| | | SAT-V | SAT-M | OSPE[a] | PVI | GPA |
| Public | 216 | 572 | 592 | 9.77 | 36.4 | 2.29 |
| Private | 97 | 547 | 564 | 8.02 | 32.5 | 1.98 |
| *F* ratio | | 8.82* | 11.11* | 8.02* | 10.24* | 14.44* |

[a] Scaled score.
* $p < .01$.

motivation, as measured by PVI. If this hypothesis is correct, then equating the two groups for these factors should result in an equalizing of GPAs.

## RESULTS

Table 3 shows the results of three analyses of covariance. The actual GPAs are shown with the *F* ratio from Table 2, so that the successive adjustments and the resulting *F* ratios may be compared. The two groups are first compared on GPA with OSPE statistically held constant. The resulting adjusted GPAs are the best estimate of GPAs if there were no differences between the two groups in intellectual factors, as measured by this test. The GPAs are then estimated for the two groups with PVI held constant. Finally, both variables are held constant. The third set of adjusted GPAs are the best estimates of what the grades would be if the groups were equated for both intellectual and nonintellectual factors.

While initially GPA for public school students was 2.29 and for private school students 1.98 with an *F* ratio of 14.44, after adjustment for both OSPE and PVI, GPA for public school students was 2.23 and for private school students 2.10 with an *F* ratio of 4.00.

These data suggest that the academic performance of the two groups as measured by GPA becomes more alike as the groups are equated on aptitude factors, motivational factors, and on both.

Finally an analysis of covariance has been computed for GPA adjusted for all four of the variables shown on Table 1. The assumption seems justified that the three measures of aptitude combined would give a better estimate of intellectual factors than a single measure.

When this adjustment is made GPAs have changed only slightly from the previous test with OSPE and PVI, even though the *F* test is no longer in the area of decision. These data are con-

TABLE 3

ANALYSES OF COVARIANCE

| GPA | Type of school | *df* | F |
|---|---|---|---|
| Before adjustment | | | |
| 1.98 | Private | 1/311 | 14.44** |
| 2.29 | Public | | |
| Adjusted for aptitude (OSPE) | | | |
| 2.04 | Private | 1/310 | 8.12** |
| 2.26 | Public | | |
| Adjusted for academic motivation (PVI) | | | |
| 2.05 | Private | 1/310 | 7.57** |
| 2.25 | Public | | |
| Adjusted for aptitude (OSPE) and academic motivation (PVI) | | | |
| 2.10 | Private | 1/309 | 4.00* |
| 2.23 | Public | | |

Note.—GPAs for public and private school students regressed for OSPE, for PVI, and for both.
* $.01 > p < .05$.
** $p < .01$.

TABLE 4

ANALYSIS OF COVARIANCE

| Adjusted GPA | *df* | *F* |
|---|---|---|
| Private = 2.22<br>Public = 2.11 | 1/307 | 2.77* |

Note.—Public and private school students regressed for OSPE, SAT-V, SAT-M, and PVI.
* $p > .05$.

TABLE 5

COMPARISON OF THE ACTUAL AND PREDICTED GPAS OF PUBLIC AND PRIVATE SCHOOL STUDENTS

| First semester GPA | Number of students | | | |
|---|---|---|---|---|
| | Public | | Private | |
| | Predicted | Actual | Predicted | Actual |
| 3.2 or higher | 25 | 25 | 3 | 3 |
| 2.8–3.19 | 37 | 35 | 8 | 10 |
| 2.40–2.79 | 33 | 37 | 12 | 8 |
| 2.00–2.39 | 48 | 41 | 25 | 32 |
| 1.60–1.99 | 41 | 42 | 18 | 17 |
| Under 1.60 | 32 | 36 | 31 | 27 |

Note.—GPA predicted from three aptitude measures and from a motivational measure (PVI).

sistent with the hypothesis that when the groups are equated for aptitude and motivation they will not differ on GPA. While one cannot conclude that there are now no differences between the groups, the groups appear to be nearly matched on performance. A further test of the data can be made by making predictions for each individual student based upon his four test scores. Table 5 shows the number of students who are predicted to achieve at the given GPA compared with the actual achievement for the 1961 group.[2] Using chi square and the predicted scores as the population parameter, the no-difference hypothesis again cannot be rejected ($p > .05$).

An additional way to determine the adequacy of the procedure is to utilize the equations from the multiple adjust-

TABLE 6

TEST SCORES AND PERFORMANCE LEVELS WITH ADJUSTED GPAS FOR THE 1960 ENTERING CLASS BY PUBLIC AND PRIVATE SCHOOL

| Type of school | *N* | Mean scores | | | | | |
|---|---|---|---|---|---|---|---|
| | | Entrance examinations | | College examinations | | Grades first semester | |
| | | SAT-V | SAT-M | OSPE [a] | PVI | GPA | Adjusted GPA [b] |
| Public | 259 | 553 | 586 | 10.49 | 37.66 | 2.32 | 2.29 |
| Private | 104 | 547 | 572 | 10.06 | 35.24 | 2.19 | 2.26 |
| Regression coefficients | | .00149 | .00023 | .03597 | .02236 | | |

[a] Scaled score.
[b] Adjusted for three aptitude measures and for academic motivation (PVI).
Best estimate of standard error of difference of means for adjusted GPA = .065.

---

[2] Because the GPA is skewed and discontinuous at some points, comparing the predicted GPA with actual GPA is difficult. The procedure followed was to calculate the predicted GPA for each student. These predicted scores were rank ordered. Since a total of 28 students (25 public and 3 private) were in the category of 3.2 or higher GPA, the 28 students with the highest predicted grade average were counted to determine how many were public school students and how many were private school students. This procedure was continued for all of the GPA categories.

ment used in Table 4 on an independent sample.

Table 6 shows the means for the class which entered 1 year prior to the class on which the analysis has been made. The regression coefficients for the 1961 sample were used to compute the adjusted means for GPA for the two groups. The accuracy of the adjustment suggests that there is some stability to the relationships found.

## REFERENCES

LATHROP, I. T., & KIEFFER, T. J. College achievement of public vs. private high-school graduates. *Clearing House,* 1959, **33**, 299–302.

MCARTHUR, C. Personalities of public and private school boys. *Harv. educ. Rev.,* 1954, **24**, 256–262.

SCHLESSER, G. E., & FINGER, J. A. *Personal Values Inventory.* Hamilton, N. Y.: Colgate University, 1961.

SELTZER, C. C. Academic success in college of public and private school students: Freshman year at Harvard. *J. Psychol.,* 1948, **25**, 419–431.

SHUEY, A. M. Academic success of public and private school students at Randolph-Macon Woman's College: I. The freshman year. *J. educ. Res.,* 1956, **49**, 481–492.

SHUEY, A. M. Academic success of public and private school students at Randolph-Macon Woman's College: II. The sophomore year. *J. educ. Res.,* 1958, **52**, 35–38.

(Received November 16, 1962)

Journal of Educational Psychology
1963, Vol. 54, No. 3, 123–131

# STUDY OF RETENTION OF SOME TOPICS OF ELEMENTARY NONMETRIC GEOMETRY[1]

ROBERT M. GAGNÉ
*American Institute for Research, Pittsburgh*

AND

OTTO C. BASSLER
*University of Maryland*

Retention was measured of the task "specifying sets, intersections of sets, and separations of sets, using points, lines, and curves," as well as of subordinate knowledges contributing to the learning of this task, in 5 groups each of 16 6th-grade students after an interval of 9 weeks. Experimental treatments during original learning represented variations in task examples and in time following practice of examples. The results show significantly lower retention for the group having the smallest variety in task examples, but no other differences. In addition, marked changes are shown between immediate and 9-week retention in the patterns of retention of subordinate knowledges.

Previous studies of the learning of mathematical knowledge have employed brief self-instructional programs to investigate some of the variables affecting achievement measured a day or two following the completion of learning (Gagné, Mayor, Garstens, & Paradise, 1962; Gagné & Paradise, 1961). In these studies, the factor of *topic sequence* appears prominently as one which exerts considerable control over the process of learning. In simple terms, it has been verified in these studies that the capability of performing certain identifiable units of subordinate knowledge creates a high probability of acquiring a new item of knowledge; whereas the lack of capability of performing any one of these same subordinate units reduces this probability to a very low value (cf. Gagné, 1962).

In a study carried out just prior to the present one (Gagné & Staff, 1963), self-instructional techniques were again used as a vehicle for the investigation of the effects of still other variables on the learning process, using materials on nonmetric geometry. The instruction was concerned with defining and specifying sets, intersections of sets, and separations of sets, using points, lines, and curves. In this study, the variables explored were (*a*) variety of examples used in the repetition of each subordinate task, and (*b*) the amount of time elapsing between the attainment of one entity of knowledge and the next. The subjects in this study were 90 sixth-grade students, randomly assigned to five experimental groups. Following the learning, a test of achievement was administered to all students.

Although the results of this previous study again confirmed the importance of the variable of topic order, no significant differences were found to be associated with the other variables of task variety and time elapsing after task practice. A number of reasons may be adduced to account for the lack of effect of these variables. Particularly, it cannot be said with confidence that any possible effects had sufficient chance to reveal themselves, in view of the fact that the average score on the final test was 18 (out of a possible 31), and thus disappointingly low.

Despite these findings, it was considered worthwhile to extend the initial study by measuring retention of knowledge in these same students several weeks later. For one thing, the retention of knowledge following learning by self-instructional means has not often been measured (cf. Stolurow, 1961). Second, it was believed possible that effects of the experimental variables might reveal themselves in retention in a more pronounced fashion than in an immediate postlearning test. Third, there was an interest in discovering not only what happened to the criterion achievement following a forgetting interval, but also what happened to the subordinate knowledges (learning sets) which have been found to contribute to this achievement during learning.

---

[1] This study is one of a series conducted by the University of Maryland Mathematics Project, and was contributed to by the staff of that Project, including J. R. Mayor, Director, Helen L. Garstens, Associate Director, Mildred B. Cole, Elise B. Cussler, and W. B. Moody, in addition to the authors.

Accordingly, an additional study was performed, as reported here, to measure the retention of the initially learned knowledge of nonmetric geometry after a 9-week interval. Measures were taken of (*a*) achievement on items matched in content with the original achievement test, and of (*b*) the performance of each of the subordinate knowledges contributing to the final task.

## METHOD

### *Learning Materials*

Details of the construction of the original learning materials are given in a previous paper (Gagné & Staff, 1963). Figure 1 indicates the topics of this material, together with their hypothesized supporting relationships. These topics were introduced in an order which took advantage of the occurrence of positive transfer to each item of knowledge represented from each subordinate item connected to it by an arrow.

### *Experimental Treatments during Learning*

Five different forms of learning programs were used during the initial learning, which occupied periods of 11 successive class days. Program El had minimal variety in its problem examples; Program E2 an intermediate variety; and Program E3 maximal variety. Program E0 contained no problem examples (except for an initial one required in the completion of each subordinate task). Program EA was the same in content as E0, but had sets of arithmetic problems inserted so as to occupy about the same amount of time as the example problems.

### *Subjects*

The subjects in the study were the sixth graders used in the previous study. Random assignment of students to each experimental treatment was made within each of the four participating classes. Because of absences, data from two subjects (in each of two different groups) could not be obtained. Accordingly, the data from two subjects, randomly chosen, were dropped from each of the other three groups. This provided a total $N$ of 80, with 16 subjects in each group, on which both original measures and retention measures were available. All of the analyses for the present study are based upon these five groups, 16 in each. A comparison of mathematics grades for these groups indicated no significant differences among them.

### *Procedure*

The procedure for administration during initial learning involved the presentation of learning programs to students in 11 booklets on 11 successive class days. Those students in Treatment E0 started later than the other students, since their material was presented in 8 booklets, rather than 11. In addition, among the various classes of students (each class containing equal numbers of students in each experimental treatment), some variations occurred in the total number of days during this learning period, owing to differences in feasibility of scheduling classes for experimental use. Accordingly, the tests of achievement which followed learning were administered to the various classes on 2 successive days during the period March 1–8, and this time represents the termination of original learning.

On the days May 9 and 10, tests of retention were administered to all groups. The retention interval for the present study may therefore be stated as approximately 9 weeks. On the first of these days, the test given was an achievement test containing items conceptually identical with those used immediately after learning. The intention was to measure not only retention of the total final task, but also of those items of subordinate knowledge which may have contributed to it.

### *Retention Measures*

The performance test given after the 9-week retention interval contained a variety of prob-

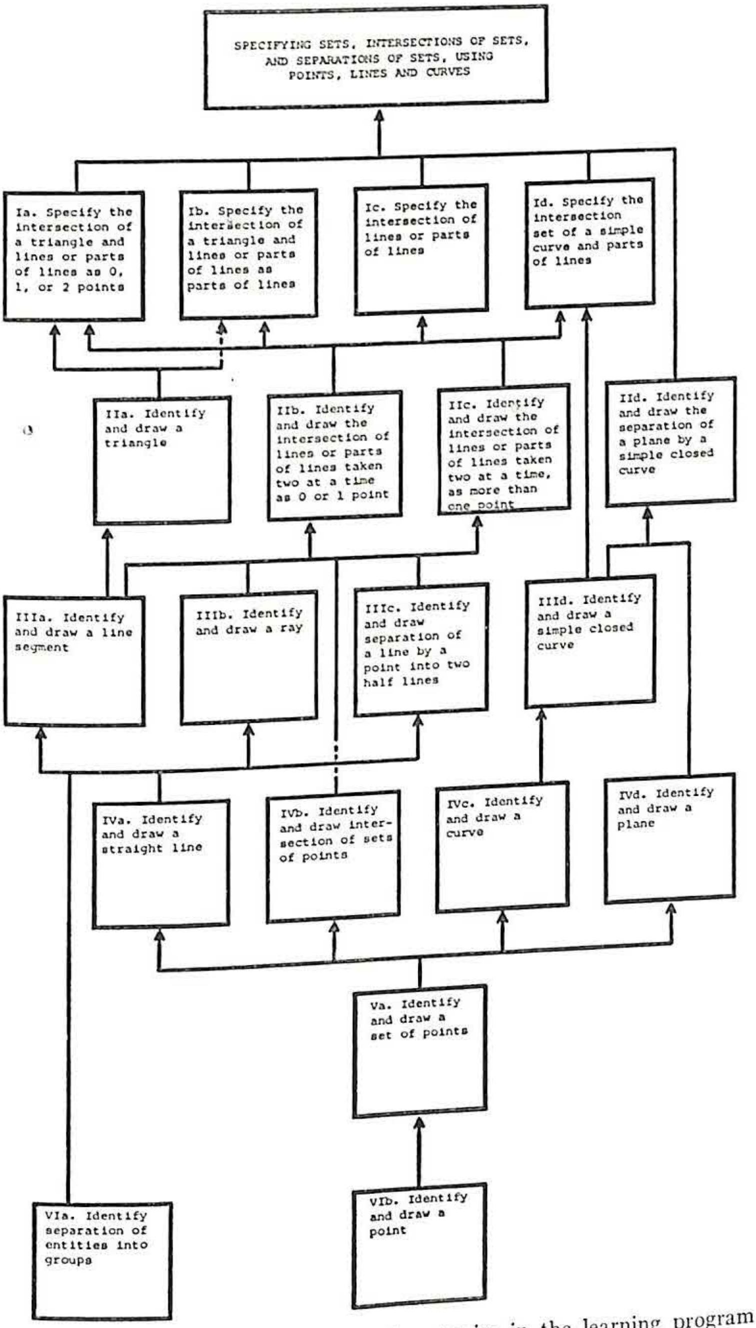

SPECIFYING SETS, INTERSECTIONS OF SETS, AND SEPARATIONS OF SETS, USING POINTS, LINES AND CURVES
Ia. Specify the intersection of a triangle and lines or parts of lines as 0, 1, or 2 points
Ib. Specify the intersection of a triangle and lines or parts of lines as parts of lines
Ic. Specify the intersection of lines or parts of lines
Id. Specify the intersection set of a simple curve and parts of lines
IIa. Identify and draw a triangle
IIb. Identify and draw the intersection of lines or parts of lines taken two at a time as 0 or 1 point
IIc. Identify and draw the intersection of lines or parts of lines taken two at a time, as more than one point
IId. Identify and draw the separation of a plane by a simple closed curve
IIIa. Identify and draw a line segment
IIIb. Identify and draw a ray
IIIc. Identify and draw separation of a line by a point into two half lines
IIId. Identify and draw a simple closed curve
IVa. Identify and draw a straight line
IVb. Identify and draw intersection of sets of points
IVc. Identify and draw a curve
IVd. Identify and draw a plane
Va. Identify and draw a set of points
VIa. Identify separation of entities into groups
VIb. Identify and draw a point

FIG. 1. Hierarchy of learning sets to be acquired as topics in the learning program employed during the original learning sessions (from Gagné & Staff, 1963).

lems requiring the specification of sets, intersections of sets, and separations of sets, using points, lines, and curves. Each item of this test was constructed by deliberately altering an equivalent item in the performance test given immediately after learning, so as to present a different context and appearance, but to maintain exactly the same idea. Thus the conceptual content of the two tests was the same, while specific item recognition was made impossible. This test was scored for a total of 20 points.

On the second testing day, a test of subordinate learning sets was administered to all subjects. Again, this test was made to have an exact correspondence with the one administered immediately after learning, in its conceptual content but not in the appearance of its items. It was designed to measure the performance of the subordinate knowledges of the hierarchy shown in Figure 1, specifically those designated as I, II, and III. Each of these subordinate knowledges was measured in two ways: (*a*) by an item of "intermediate" difference in context from that of the task as originally presented during learning, and (*b*) by an item of great difference in context from the originally presented learning task. (In the preceding learning study, three degrees of contextual difference were employed, including one of "minimal" difference in context.) As it turned out, all results obtained with measures based on items of intermediate contextual difference were essentially the same as those obtained with items of large contextual difference. Accordingly, only the former set of measures is reported here, in the interest of simplicity of presentation. For this measure of retention of subordinate learning sets, scoring was pass or fail for each learning set at each degree of contextual difference.

## RESULTS

The results of the study may be considered as providing evidence in four major categories, as follows:

1. To what extent are performances immediately following learning correlated with performances measured after a 9-week interval?
2. How much retention was there?
3. What were the effects on retention of experimental treatments during learning?
4. As compared to forgetting of final performances established during learning, what happens to the retention of *subordinate* knowledges?

### *Correlations of Measures of Immediate and 9-Week Retention*

For the total group of 80 subjects, the product-moment correlation between achievement scores immediately after learning and those attained after a 9-week retention interval was .62. This is a moderate relationship, and needs to be interpreted in the light of the fact that the reliability of the achievement test (according to data of the original study based upon 90 subjects) was found to be only .64, an uncorrected correlation for odd-even items.

In contrast to this finding are the product-moment correlations of .41 between scores on subordinate learning sets (of intermediate difference in context) measured immediately and after 9 weeks; and .46 between scores on learning sets of great difference in context measured on these two occasions. These are low correlations, obtained on measures which have uncorrected reliabilities of .90, by odd-even correlations obtained in the original group.

These results may therefore be said to reveal moderately good consistency between initial and 9-week measures on the final task, but a relatively low consistency between measures of subordinate knowledges on these two occasions. The implication is that during a period of potential forgetting, something different may be happening to the retention of a total performance and to the retention of subordinate knowledges which support this performance during original learning. A further analysis of this effect is made in a subsequent section.

### *Amount of Retention*

Retention of the final task was measured by means of a 20-point achievement test which, as has previously been mentioned, contained items identical in concept and different in appearance from items used to measure immediate postlearning achievement. The retention of the 12 subordinate learning sets at Levels I through III (see Figure 1) was measured by a similar procedure. The means and standard deviations of these measures for both occasions, and for the five groups of the experiment, are given in Table 1.

As can be seen from an examination of Table 1, retention as measured by scores on the achievement test was very

TABLE 1

MEANS AND STANDARD DEVIATIONS OF ACHIEVEMENT SCORES AND SUBORDINATE LEARNING SET SCORES ATTAINED IMMEDIATELY FOLLOWING LEARNING, AND AFTER 9 WEEKS, FOR THE FIVE EXPERIMENTAL GROUPS, TOGETHER WITH RETENTION RATIOS
($N = 16$ in each group)

| Measure | Experimental group | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | E1 | | E2 | | E3 | | EA | | E0 | |
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Achievement (immediate) | 11.2 | 2.4 | 10.8 | 3.5 | 9.7 | 3.7 | 10.4 | 4.0 | 10.0 | 3.8 |
| Achievement (9-week retention) | 7.9 | 2.2 | 11.3 | 2.9 | 10.9 | 3.2 | 10.3 | 3.3 | 12.4 | 3.7 |
| Retention ratio (achievement) | .75 | | 1.12 | | 1.20 | | 1.08 | | 1.28 | |
| Learning sets (immediate) | 8.8 | 3.3 | 8.4 | 3.0 | 9.2 | 2.9 | 8.9 | 3.3 | 9.2 | 3.0 |
| Learning sets (9-week retention) | 4.9 | 1.9 | 6.4 | 1.3 | 6.1 | 1.5 | 6.1 | 2.1 | 6.4 | 2.5 |
| Retention ratio (learning sets) | .60 | | .88 | | .80 | | .79 | | .70 | |

high in all groups except E1 (the group having a narrow variety of examples during learning). In fact, as the ratios indicate, retention in the other four groups ranged from 108% to 128%. It must of course be remembered that the test of achievement given immediately following learning constitutes an additional practice period. Consequently, these heightened achievement scores, obtained after a 9-week interval, cannot be interpreted as uncontaminated evidence of *reminiscence* (cf. McGeoch & Irion, 1952, p. 162). Nevertheless, these scores do constitute good evidence that this kind of material, when learned by means of a carefully constructed instructional program, is *highly resistant to forgetting.*

Somewhat in contrast to these findings are those pertaining to the retention of subordinate learning sets; their losses through forgetting appear to be considerably greater. As the ratios indicate, retention of numbers of learning sets ranges from 60% to 88% in the five experimental groups. As shown in the previous study (Gagné & Staff, 1963), these subordinate learning sets are of great importance to the achievement of the final task. But the present evidence suggests that their importance diminishes once the final task has been achieved. Specifically, forgetting of subordinate learning sets may occur independently of, and without effect upon, retention of the total task which has been achieved through learning.

### *Effects of Experimental Treatments*

The results given in Table 1 also indicate the effects on retention scores of experimental treatments applied during the original learning sessions. The variances of scores for these groups are found to display nonsignificant differences. But the means for achievement scores attained after 9 weeks differ significantly, when tested by means of an analysis of variance technique ($F = 4.46$, $df = 4/75$, $p < .01$). Applying a test to the differences of individual means (Dixon & Massey, 1957, p. 152), it is found that the mean of E1 differs significantly from that of E0 at the 1% level, while other comparisons show no significant differences. Also at the 1% level, the mean of E1 differs from the combined mean for the other groups. In other words, the only outstanding difference among these means is that exhibited by Group E1, which is lower than all the others.

As for the retention of subordinate learning sets, these means exhibit no significant differences among them. An analysis of variance applied to these values yielded $F=1.85$, $df=4/75$, $p > .10$. Thus, the low retention scores of group E1 occur despite the fact that this group has retained subordinate learning sets not significantly less well than the other groups of the experiment.

### *Retention of Subordinate Knowledges*

As results previously described have shown, retention of the capability of performing a final task appears to be better than retention of the subordinate learning sets which originally contribute to the learning of that capability. Some of these learning sets are forgotten, even though performance of the total task remains at a level as good as that originally attained. Are there differences in the *pattern* of retention for these learning sets immediately after learning and 9 weeks later? For example, do those at Level III (see Figure 1) tend to be forgotten, while those at Level I tend to remain, or vice versa?

As previous studies have shown (Gagné et al., 1962; Gagné & Paradise, 1961) the pattern of learning sets which an individual learner has attained displays a remarkably regular pattern, when measured immediately following learning. It is a pattern which implies that the learner continues to acquire learning sets at higher and higher levels of the hierarchy until he reaches a particular point, beyond which he acquires no more at all. (It is as if he "drops out" of the learning enterprise at this point.) Findings in the immediately previous study (Gagné & Staff, 1963) were entirely consistent with this general result.

Examination of individual patterns of learning set *retention*, however, shows that these patterns are strikingly different. The forgetting of individual learning sets appears to be irregular, or random. Learning sets at one level do not appear to be forgotten with greater frequency than those at another level. An illustration of this effect is given for a typical individual learner in Figure 2.

This irregularity of pattern may be measured as the *number of discontinuities* which occur in the score (pass or fail) of an individual on learning sets when these are arranged in the order IIIa, IIIb, IIIc, IIId, IIa, IIb, IIc, IId, Ia, Ib, Ic, Id. For an individual who passes all of them, the discontinuities will of course be 0. For a student who begins to fail learning sets at some particular level, ideally expected numbers of discontinuities would be 1. A score higher than this indicates by its size the irregularity of pattern exhibited in the attainment of learning sets. For the student whose scores are shown in Figure 2, for example, the *discontinuity score* for learning sets measured immediately after learning is a not atypical 1; but the score for the same learning sets measured after 9 weeks is 7.

Such a discontinuity score was counted up for each subject in each

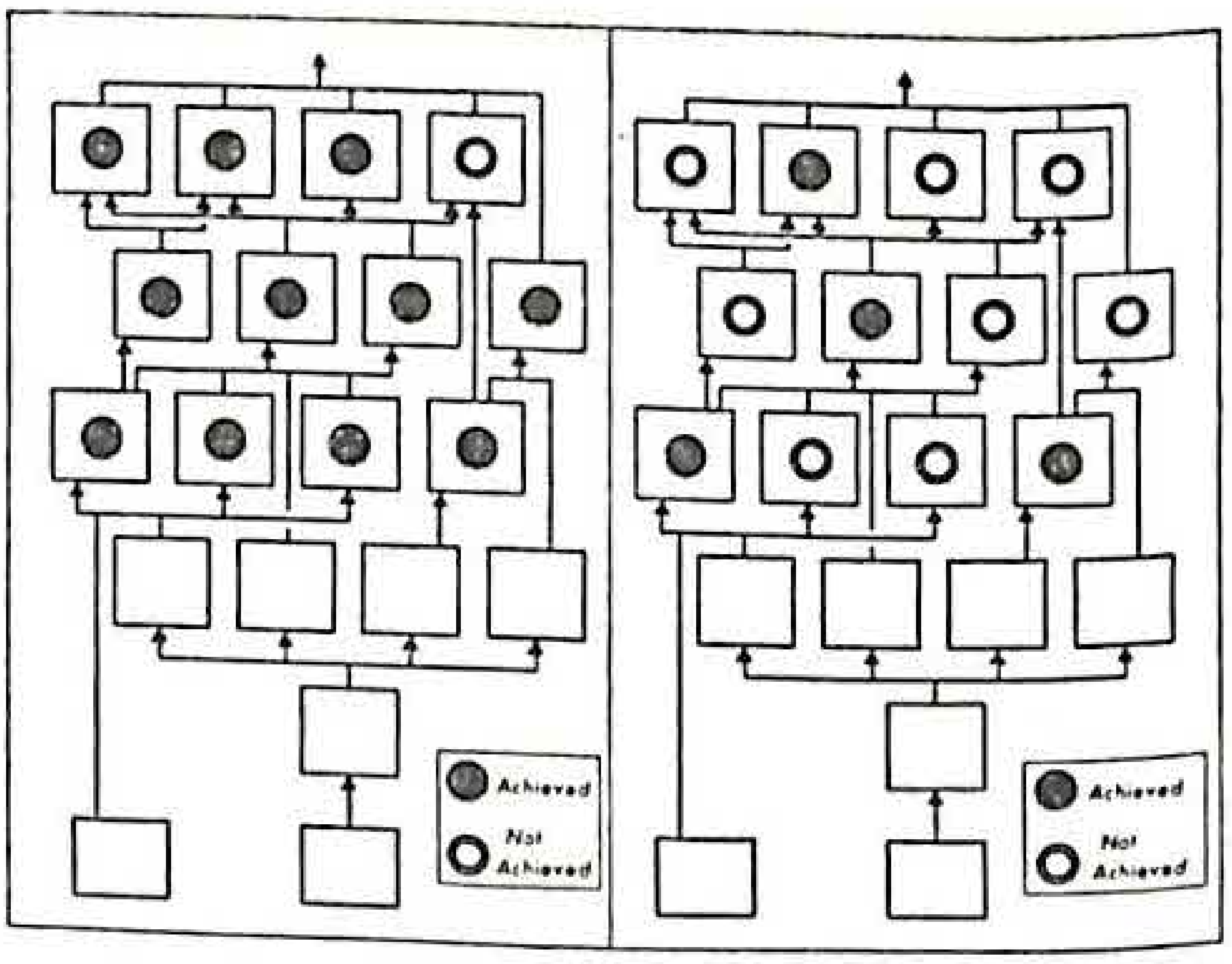


Fig. 2. Comparison of the pattern of learning set achievement exhibited by one subject (JO) when measured immediately following learning (left figure) and on a test of retention 9 weeks later (right figure).

TABLE 2

MEANS AND STANDARD DEVIATIONS OF DISCONTINUITY SCORES ON THE PASSING AND FAILING OF LEARNING SET ITEMS ARRANGED IN A NUMERICAL ORDER, MEASURED IMMEDIATELY AFTER LEARNING AND 9 WEEKS LATER, FOR THE FIVE EXPERIMENTAL GROUPS ($N = 16$ in each group)

| | Experimental group | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | E1 | | E2 | | E3 | | EA | | E0 | |
| Learning set measure | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Immediate | 1.1 | 1.1 | 1.1 | .6 | 1.4 | .6 | 1.3 | 1.0 | 1.4 | 1.0 |
| After 9 weeks | 4.8 | 1.7 | 5.0 | 1.6 | 4.8 | 1.2 | 4.9 | 1.5 | 4.8 | 1.7 |

experimental group. The means and standard deviations of these scores are presented in Table 2.

As the table clearly indicates, the mean discontinuity scores for learning sets measured immediately after learning are markedly less than those measured after 9 weeks. When the differences between these means are tested using the $t$ test (for correlated measures), all five turn out to be significant beyond the 1% level of confidence. Thus all of the data on retention of subordinate learning sets support the conclusion that forgetting strikes them in an irregular and apparently random fashion. Despite the fact that the final task in four of these groups exhibits more than 100% retention, subordinate knowledges have been forgotten in an irregular manner.

## DISCUSSION

The results have shown that retention of the final task of "specifying sets, intersections of sets, and separations of sets, using points, lines, and curves" is extremely high after 9 weeks during which no recall was encouraged or required within the school setting. The measured retention was more than 100% for four of the five experimental groups of the experiment, and 71% in the fifth group. Of course, this measure is in all cases affected by the practice provided by the test used to measure achievement immediately following the original learning. It is reasonable to characterize the retention as remarkably good.

The finding that, in one group, retention of the final task was *lower* than all the rest deserves further comment. This experimental group (E1) originally learned by means of an instructional program which contained several examples of *minimal variety* following the attainment of each learning set (i.e., topic) in the program. The results therefore show that, so far as retention is concerned, such a treatment is less effective than that used for Group E2, containing examples of intermediate variety, as well as that for E3, with examples of maximal variety. But the remarkable thing is that the "minimal variety" of examples used for E1 seems also to have been less effective than *no examples* at all, which was the treatment accorded Groups EO and EA. (It will be recalled that "no examples" in this context means none beyond the original presentation of the subordinate task. Actually, each task of the hierarchy occurred at least once during the learning, in all groups; "examples" means those instances of the task beyond this initial one.) It therefore appears from the present results that the presentation of additional examples of minimal variety has had an *actively deleterious* effect upon retention.

This particular finding is not an easy one to account for, in terms of hypothesized mechanisms of forgetting. It is possible, however, that any treatment which tends to prevent the generalization of subordinate knowledges has an adverse effect upon their later recall and use. The subordinate learning sets in a hierarchy (such as that of Figure 1) are, after all, supposed to operate in both learning and retention by continually generalizing to new situations, whether these are in the "higher level" learning sets or in the execution of a final task. It is conceivable that repetition of these subordinate tasks within a narrow context serves to reduce their generalizability, as is true with simpler habits (LaBerge, 1961). Such a narrowing of generalizability might reduce the effectiveness of these learning sets to a point even below that attained by learning not attended by any additional "example practice," as in Groups EO and EA.

We do not feel entirely confident of this finding, and even less so of our proffered explanation. A replication should certainly be carried out, and if it is verified, additional studies need to be undertaken to explore thoroughly the limits of the effect of "narrow repetition" on retention of meaningful subject matter.

The remaining finding of considerable significance in the present study is the differential, random seeming, forgetting of subordinate knowledges over the 9-week period, occurring at the same time as high retention of a capability for performing the final task. It has often been supposed that learning sets, in such an instructional sequence as this, may have a positive function in learning which is not displayed (or required) in the final performance. The subordinate learning sets, it has been said (Gagné & Paradise, 1961), "mediate positive transfer of training" to higher level learning sets. Once they have performed this function, however, they may suffer forgetting, or "drop out." Apparently, this is what happened to many of the learning sets included in the instructional sequence of the present experiment. As opposed to the building-block character they have when measured immediately after learning, they disappear in the irregular fashion depicted in Figure 2.

What are the consequences of the forgetting of these subordinate knowledges? If one is dealing only with a particular capability, such as that provided in the final task here employed, there are no adverse consequences; the student can still perform capably on the final task. However, it is an important consideration that these same subordinate knowledges may have to perform this learning function *again*, when the student is asked to learn a *new* task. And if they are not recalled at that time, as our previous studies have shown, learning will not occur.

Forgetting is a process which takes place all the time. The forgetting of those subordinate knowledges which may be needed in the learning of a new task is something which probably should be expected to happen in a more or less unpredictable way. The variables used in the present experiment (increasing variety of examples, provision of time following example practice) have not been shown to be effective in preventing forgetting. Taken as a whole, therefore, the results emphasize the practical importance of *review* in any sequence of instruction. An instructional program accomplishes this by requiring systematic recall of previously learned knowledge before asking the student to achieve a new item of knowledge. In this somewhat indirect way, the importance of the "recall item" (cf. Gagné, 1962) for effective learning is again emphasized.

Our experiment provides no hint as to how some learning sets come to be retained in any near permanent fashion. A best guess would be that such retention is fostered not primarily by "variety of learning context" (as used in the present experiment) but rather by "variety of use." If a given learning set is used to support the learning of a great variety of higher-level knowledges, perhaps this kind of variety makes it more recoverable and less subject to forgetting. This is an interesting speculation, which cannot, however, be supported by any currently available evidence.

## SUMMARY

A study was carried out to measure the retention of knowledge described by the task of "specifying sets, intersections of sets, and separations of sets, using points, lines, and curves," as well as of subordinate knowledges (learning sets) contributing to the original learning of this task. During the initial learning period, five groups of sixth-grade students, 16 in each group, worked through an instructional program having the same topical content, but constructed so as to constitute five different experimental treatments. Group El practiced examples of each subordinate task which had minimal variety in context; Group E2, intermediate variety; Group E3, great variety; Group EO, no examples; and Group EA, unrelated tasks (arithmetic) in place of examples.

Retention was measured after a period of approximately 9 weeks. The results were as follows:

1. The mean retention score of Group El was significantly lower than that of the other groups, whose means did not differ from each other. This evidence suggests that "narrow practice" of examples has a negative effect on retention. At the same time, no advantage in retention was found for increased variety of task examples.

2. Except for Group El, retention was more than 100% by comparison of scores on a test given immediately following learning and on a test with items of identical conceptual structure given 9 weeks later. (Item appearance was entirely different on the two occasions.) It is pointed out that the first test constituted a practice trial, and that true "reminiscence" cannot therefore be said to have occurred. But retention in these four groups was very high.

3. In contrast with achievement scores, the percentage of retention of subordinate learning sets ranged only from 60% to 88%, indicating considerably greater losses. In addition, the pattern of forgetting revealed by the retention test of learning sets was found to be an irregular one in contrast to the pattern of these subordinate knowledges exhibited immediately after learning. These results are interpreted as throwing light upon the manner of "dropout" of subordinate knowledges, and emphasize the importance of brief "recall items" in a sequence of instruction for acquiring new knowledge.

## REFERENCES

DIXON, W. J., & MASSEY, F. J., JR. *Introduction of statistical analysis.* (2nd ed.) New York: McGraw-Hill, 1957.

GAGNÉ, R. M. The acquisition of knowledge. *Psychol. Rev.*, 1962, **69**, 355–365.

GAGNÉ, R. M., MAYOR, J. R., GARSTENS, H. L., & PARADISE, N. E. Factors in acquiring knowledge of a mathematical task. *Psychol. Monogr.*, 1962, **76**(7, Whole No. 526).

GAGNÉ, R. M., & PARADISE, N. E. Abilities and learning sets in knowledge acquisition. *Psychol. Monogr.*, 1961, **75**(14, Whole No. 518).

GAGNÉ, R. M., & STAFF, University of Maryland Mathematics Project. Some factors in learning non-metric geometry. *Monogr. Soc. Res. Child Develpm.*, 1963, in press.

LABERGE, D. Generalization gradients in a discrimination situation. *J. exp. Psychol.*, 1961, **62**, 88–94.

MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning.* New York: Longmans, Green, 1952.

STOLUROW, L. M. *Teaching by machine.* (Cooperative Research Monograph No. 6.) Washington, D. C.: United States Department of Health, Education, and Welfare, Office of Education, 1961.

(Received January 23, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 3, 132–137

# DIFFERENCES IN PARENT AND CHILD REPORTS OF THE CHILD'S EXPERIENCE IN THE USE OF MONEY

HELEN R. MARSHALL[1]
*Drexel Institute of Technology*

This investigation of the meaning of differences in parent and child reports of the child's experience analyzed differences in reports about recent use of money by 180 children aged 10–12 years. The results suggested that parent and child reports are not similar enough to be interchangeable, that differences between reports relate to child characteristics, and that child reports are more closely related to other variables than are parent reports. A complementary study of interrelations of the various measures indicated that children's financial knowledge does not differ from other knowledge in its association with IQ of the child.

Research about home determinants of children's behavior is plagued by many problems of measurement. Among these is a lack of knowledge of the meaning of differences in parent and child reports of the child's experience. This investigation explored this problem through analysis of differences in parent and child reports of the child's use of money in the past week and year. Such home experiences appeared to lack emotional overtones leading to gross distortions, and to be of enough interest to both parents and children to be remembered.

Exploration of differences focused on three questions:

1. How much do parent and child reports agree and disagree?

2. Do characteristics of the child contribute to differences?

3. Whose reports bear closer relationships with other variables?

A secondary purpose was to study interrelations of measures used in the analysis of differences, i.e., parent and child reports of the child's use of money, children's tested characteristics, and children's scores on tests of financial knowledge and responsibility. This phase of the study centered on testing the hypothesis that children's financial knowledge does not differ from other knowledge in its association with IQ of the child. Relations between children's financial knowledge and both home experiences and some child characteristics were investigated by Marshall and Magruder (1960), but the influence of IQ was not explored in that study.

## METHOD

The subjects were 180 children and one parent of each who resided in the same agricultural region of Kentucky, although in different counties, as the children studied by Marshall and Magruder. The sample consisted of 30 boys and 30 girls at each of the ages of 10, 11, and 12 years, selected from 716 children tested at county consolidated schools to fill an analysis of variance design for another study (Marshall, 1963).

The sample included all levels of socioeconomic status, as determined by the Cornell 15-point scale (Danley & Ramsey, 1959). The mean IQ from the California Test of Mental

---

[1] The investigation reported in this paper was conducted in part in connection with a project of the Home Economics Department of the Kentucky Agricultural Experiment Station and is published with permission of the Station Director. It was written after the author was employed elsewhere.

Maturity (1957 edition) was 102, and the standard deviation was 15.

All group and individual tests and questions were administered to the children at school in three sessions within a period of 2 weeks by the investigator and assistants. Parents were interviewed at home within the following month.

The following child experiences were included in the study of differences in parent and child reports.

1. Experience with Money test (Marshall & Magruder, 1960). A high score on this test indicates a wide variety of experience.
2. Past week's wages—the amount earned by the child at home and elsewhere.
3. Past year's wages—the sum of amounts reported paid for any job performance.
4. Past year's gifts of money from any source, exclusive of allowances.
5. Past week's spending money from any source.
6. Number of expenses assigned to the child's spending money. The number of 14 spending categories (food snacks, toys, clothing, gifts, school needs, dues, etc.) that both parent and child stated had to be purchased, if at all, from the child's spending money.

*Difference scores* for the first four experiences were determined by subtracting the parent's report from the child's, and adding a constant to make all scores positive. The mean difference between three reports of the past week's spending money was used as the difference score for the fifth experience, and was used also as an estimate of variability of spending money. Two of these three reports were from interviews with the child and parent, and the other was written on the child's initial information sheet. The number of specific expenses that both parent and child indicated were assigned to spending money was used as an *agreement score* for the fifth experience.

Characteristics of children included in the study of interrelations were chronological age, IQ on the California Test of Intelligence, score on the Children's Manifest Anxiety *CMA* scale Castaneda, McCandless, & Palermo, 1956), number of siblings, and family socioeconomic status on the Cornell scale.

The seven estimates of children's financial knowledge and responsibility employed were as follows: (*a*) The Financial Knowledge Test was a 13 multiple-choice question test developed for this study. Questions were concerned with family money management, and with relative costs of food, clothing, and cars. (*b*) The Coin test estimated accuracy in pricing inexpensive toys, as described by Marshall and Magruder (1960). (*c*) The numerical reasoning score was obtained from a subscale of the California Intelligence Test. (*d*) The Harris scale concerned "behavior classifiable as reliable, accountable, loyal, or doing an effective job" (Harris, 1957, p. 322). (*e*) The Lie scale included in the *CMA* scale was used as an estimate of irresponsibility, on the assumption that realistic description of self is an aspect of "accountable" behavior. (*f* and *g*) Parent judgments of responsibility for money and work compared the child to siblings and the child's classmates known by the parent.

## RESULTS

Parent and child reports of the same child experience had a slight resemblance, as is shown in Table 1. These small *r*'s do not support assumptions that the reports are identical, interchangeable, or reports about the same experience.

The *r*'s between two reports of the same experience contrast sharply with *r* of .87 obtained between parent and child reports of home furnishings and equipment on the Cornell Scale of Socioeconomic Status in this sample. Agreement about possessions was significantly greater than agreement about an experience of the child.

One reason for the difference in parent and child reports is suggested by their mean report scores shown in Table 1. Children said they had more experience using money than their parents reported, and this difference was significant at the .01 level for all topics except the Experience with Money score.

The size of the difference between child and parent reports was extremely variable and often very large. For the several reports of amounts of money, the range of differences was more than eight times the size of the mean of either child or parent reports.

Parents and children did not differ consistently over the five experience topics listed in Table 1. Intercorrela-

TABLE 1

Correlations between and Means of 180 Child and Parent Reports of the Child's Experience

| Topic of report | *r* | Mean of reports by Children | Mean of reports by Parents |
|---|---|---|---|
| Experience with Money score | .33** | 11.7 | 11.3 |
| Past week's wages | .07 | $ .37 | $ .27 |
| Past year's wages | —[a] | —[a] | —[a] |
| Boys | .44** | $46.36 | $39.01 |
| Girls | .16* | $23.46 | $11.54 |
| Past week's spending money | | | |
| Interview week | .37** | $ .69 | $ .60 |
| Information sheet week | .18* | $ 1.06 | |
| Number of assigned expenses | .20* | 4.8 | 3.9 |

[a] Sex difference in *r*'s and between means significant at .05 level.
* $p \leq .05$.
** $p \leq .01$.

tions of these difference scores failed to exceed ± .14 in size.

The possibility that child characteristics contributed to differences in reports was explored by correlating difference scores with the child characteristics listed in the first column of Table 2. Significant *r*'s were obtained for chonological age, IQ, the two money knowledge tests, and the Harris Social Attitudes Scale. As scores on these child characteristics increased, the child and parent agreed on more assigned expenses (*r*'s from .17 to .28), and the child's estimate differed less from the parent's report of variety of experiences and of the past year's gifts (from −.17 to −.27). These correlations suggest that older, brighter, more knowledgeable, and more responsible children differed less from their parents in reporting some topics than their less favored age peers.

Significant *r*'s for difference scores involved the same characteristics and test scores as those shown for child reports in Table 2. Correlations involving parent reports, shown also in Table 2, were less often significant. Difference scores and child reports varied in accordance with the child's other characteristics, while parent reports often did not. In short, child reports were more closely related to other variables than parent reports of the same experience. Child reports had 19 significant *r*'s with child characteristics and knowledge, exclusive of relations for variability of spending money, and those involving sex differences, in contrast to 8 significant *r*'s involving parent reports.

The second part of this investigation was a study of interrelations of measures used in the preceding analysis of differences, particularly in regard to their support for the hypothesis that children's financial knowledge does not differ from other knowledge in its association with IQ of the child.

The obtained correlations support this hypothesis, and also provide evidence that IQ had closer relations with children's financial knowledge and responsibility than any other factor included in this study. A comparison of *r*'s involving IQ with *r*'s involving other variables in Tables 2 and 3 makes this clear. For example, correlations between IQ and determinations of children's financial knowledge and responsibility,

TABLE 2

CORRELATION OF CHILD AND PARENT REPORTS OF THE CHILD'S EXPERIENCE WITH ESTIMATES OF CHILDREN'S CHARACTERISTICS, FINANCIAL KNOWLEDGE, AND RESPONSIBILITY

| Estimates | Child's experience as reported by child (C) and parent (P) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Experience with money | | Number of assigned expenses | | Past week's spending money | | Past week's wages | | Past year's wages | |
| | C | P | C | P | C | P | C | P | C | P |
| CA | .09 | .12 | —[a] | —[a] | .30** | .20** | —[a] | .02 | .15* | .08 |
| IQ | .21** | .13 | .21** | .18* | .07 | .04 | −.13 | .07 | −.07 | .14 |
| *CMA* scale | −.01 | −.04 | −.18* | −.05 | −.01 | −.01 | −.01 | −.04 | .11 | .06 |
| Number of siblings | −.19* | −.11 | −.05 | −.02 | −.08 | −.13 | −.09 | −.08 | −.08 | −.02 |
| Socioeconomic status | .19* | .16* | .20** | .11 | .29** | .08 | .07 | .12 | .27** | .15* |
| Financial knowledge | .06 | .05 | .18* | .06 | .20** | .12 | .08 | −.10 | .03 | .06 |
| Coin test | .09 | .11 | −.03 | −.07 | .23** | .06 | .02 | −.10 | −.05 | .00 |
| Numerical reasoning | .10 | .08 | .23** | .17* | .11 | −.05 | .05 | .08 | .04 | .22** |
| Harris scale | .04 | .03 | —[a] | .14 | .05 | .05 | .06 | −.02 | −.02 | .08 |
| Lie scale | −.15* | −.08 | −.18* | −.14 | −.14 | −.09 | −.16* | −.05 | −.13 | −.06 |
| Responsibility for money | .15* | .11 | —[a] | .08 | .30** | .22** | .05 | .12 | .11 | .17* |
| Responsibility for work | −.03 | .03 | .10 | −.07 | −.12 | .02 | .00 | .08 | −.01 | .04 |

[a] Sex difference in *r*'s significant at .05 level. In all instances, *r*'s for boys were positive (from .17 to .43), but *r*'s for girls were negative (from —.03 to —.17).
* $p \leq .05$.
** $p \leq .01$.

shown in Table 3, were larger and more often significant than those obtained in similar relations for the characteristics of chronological age, manifest anxiety, number of siblings, and socioeconomic status. Also, these correlations were larger and more often significant than correlations between the seven knowledge and responsibility scores and either child reports of experience or the similar reports by parents, shown in Table 2.

Several aspects of experience with money were associated with financial knowledge and responsibility in ways not accounted for by IQ. Children's reports of experience with money had closer relations with socioeconomic status and chronological age than with IQ, as is shown in Table 2.

Obtained correlations suggest that three aspects of the child's experience with money relate to children's financial knowledge and responsibility.

1. The number of expenses the child stated were assigned to his spending money correlated positively with knowledge of money and negatively with the Lie scale, as is shown in Table 2. This aspect of experience also had positive relations with determinations of responsibility for boys, but not girls.

2. Receiving more weekly spending money accompanied more knowledge and a more favorable judgment about money responsibility from the child's parents, as is shown in Table 2. These positive relations were not the consequence of differences associated with op-

TABLE 3

CORRELATION OF CHARACTERISTICS OF CHILDREN WITH THEIR FINANCIAL KNOWLEDGE AND RESPONSIBILITY SCORES

| Knowledge and responsibility scores | CA | IQ | *CMA* scale | Number of siblings | Socio-economic status |
|---|---|---|---|---|---|
| Financial Knowledge Test | .31** | .44** | −.08 | −.15* | .15* |
| Coin Test of Knowledge of Money | .21** | .21** | −.02 | −.14 | .09 |
| Numerical reasoning | −.06 | .70** | −.22** | −.15* | .16* |
| Harris Social Attitudes Scale | .18* | .47** | −.35** | −.18* | .10 |
| Lie scale | −.24** | —[a] | −.15* | −.02 | −.11 |
| Parent judgment money responsibility | .09 | .15* | −.01 | −.11 | —[a] |
| Parent judgment work responsibility | −.04 | .09 | −.03 | −.15* | .17* |

[a] Sex difference in *r*'s significant at .05 level. Between IQ and the Lie scale, the *r* for boys was −.44, and was .02 for girls. The *r* between socioeconomic status and money responsibility was .32 for boys and −.19 for girls.

* $p \leq .05$.

** $p \leq .01$.

portunities to earn money. Wage amounts failed to relate to financial knowledge and responsibility.

3. Variability in the weekly amount of the child's spending money correlated positively with financial knowledge (.17) and the Coin test (.18). However, variability of amount had an *r* of .74 with the total amount received in 3 weeks and may not be a separate factor.

Relations for these three aspects of experience give added details and support for the conclusion by Marshall and Magruder (1960) that children's financial knowledge increases as the child's spending money increases.

## DISCUSSION AND CONCLUSIONS

The evidence of this investigation suggests that research use of either the child or parent report of the child's experience will have the consequence of different results than would be obtained with use of the other. Use of the child report seems to increase the number of relations obtained for home determinants of behavior, while use of the parent report could eliminate relations involving personality factors that contribute to differences in child behavior.

There was a suggestion of a difference in the importance of reports to parents and children. All child interviewers spontaneously described the children's delight in the opportunity to describe their experiences. When asked, parent interviewers said, "Parents attempted to give full and honest reports." This possibility could explain the greater detail found in child reports. Perhaps a difference in the importance of situations to parents and children will be a cause of differences in reports of any experience.

In this investigation of the meaning of differences between parent and child reports of the child's use of money, the results suggested that:

1. Child and parent reports of the same experience of the child are not similar enough to be interchangeable.

2. Parents and children do not differ consistently in their reports about several experiences.

3. The size of the difference is influenced by general characteristics of the child. In general, as "desired" characteristics of the child increase, the size of the difference between reports decreases.

4. Child reports vary with the child's abilities, while parent reports often did not.

5. Child reports have closer relations with other variables than parent reports.

The analyses of intercorrelations of measures indicated that children's financial knowledge does not differ from other knowledge in its association with IQ of the child. The results support two conclusions: (*a*) IQ relates more closely to children's financial knowledge and responsibility than any other factor studied for these 10–12 year old children, and (*b*) assignment of specific expenses to spending money and the use of varying and relatively large amounts of spending money had closer relations with financial knowledge and responsibility than the other experiences included in children's reports.

## REFERENCES

CASTANEDA, A., MCCANDLESS, B. R., & PALERMO, D. S. The children's form of the manifest anxiety scale. *Child Develpm.*, 1956, **27**, 317–326.

DANLEY, R. A., & RAMSEY, C. E. Standardization and application of a level-of-living scale for farm and nonfarm families. *Cornell U. Agric. Exp. Sta. Memoir,* 1959, No. 362.

HARRIS, D. B. A scale for measuring attitudes of social responsibility in children. *J. abnorm. soc. Psychol.,* 1957, **55**, 322–326.

MARSHALL, HELEN R. The relation of giving children an allowance to children's money knowledge and responsibility and to other practices of parents. *J. genet. Psychol.,* 1963, in press.

MARSHALL, HELEN R., & MAGRUDER, LUCILLE. Relations between parent money education practices and children's knowledge and use of money. *Child Develpm.,* 1960, **31**, 253–284.

(Received January 17, 1962)

*Journal of Educational Psychology*
1963, Vol. 54, No. 3, 138–143

# ITEM SCRAMBLING IN A SELF-INSTRUCTIONAL PROGRAM[1]

GERALD R. LEVIN AND BRUCE L. BAKER

*Brown University*

A 4-unit, self-instructional program on geometry was used to study the effect of presenting program items in standard or random order. 36 2nd graders, divided into 2 matched groups, went through 2 versions of the program which differed only with respect to the sequence of items in the 3rd unit. The control group went through those items in the standard order; the experimental group in a random sequence. No statistically significant differences between the groups were found on indices of acquisition, retention, or transfer. The generality of the findings was discussed in light of program effectiveness and the correlation of IQ and performance.

Decisions about the sequence in which material is presented are inherent in the process of constructing self-instructional programs. In a program designed to teach several concepts, one concept must be selected for initial treatment. If 10 items are used to teach a concept, these items must be ordered. While every programmer must make assumptions about ordering topics and items, there is little experimental evidence available which bears on the validity of such assumptions.

Gavurin and Donahue (1961) have conducted an experiment directly relevant to the question of item sequence. They used the first 29 items of a program on psychology (Holland & Skinner, 1961) and presented them in a standard sequence to one group of adults, and to another group in a random sequence. Their method of scrambling items was a combination of what might be termed *block* scrambling and *item* scrambling. The 29 items were divided into blocks of 10, 10, and 9 items. The order of blocks was scrambled for the experimental group. Within each block, the 9 or 10 items also were scrambled so that the final random order contained more local homogeneity than would have been expected if scrambling had taken place within the total pool of 29 items. The subjects were required to reach a criterion of an errorless trial on all 9 or 10 items within a block before moving on to the next block. It was found that the group which went through the items in a random sequence required significantly more trials to criterion and made more errors during acquisition. However, when retention was tested a month later, the superiority of the group which went through the standard sequence of items was not statistically significant. Unfortunately, the design of their study did not permit a comparison between the two treatments with the number of trials held constant.

The present experiment, like the study by Gavurin and Donahue, was conducted to determine the importance of presenting items in a standard, logical

---

[1] The research reported herein was supported in part by Grant No. 7-07-088.00 from the United States Office of Education, Department of Health, Education, and Welfare.

The authors are indebted to E. Fitzgerald, Superintendent of the Bristol, Rhode Island, School Department for his cooperation in facilitating this research. A. Hilton's and N. Maurer's assistance in the collection and analysis of data is gratefully acknowledged.

sequence which had been arrived at on the basis of prior planning and experimentation. The original impetus to the study was a concern that if a program were eventually to be evaluated in terms of criteria of retention and transfer, then error rate during acquisition might not be a satisfactory criterion to use while developing a program. Accordingly, the random presentation of program items appeared to be a convenient means of varying error rate during acquisition while holding other program characteristics constant. The purpose of the study, then, was to determine what happended to varied criteria of learning when item sequence was varied in a portion of a self-instructional program.

## METHOD

A matched-group experiment was planned which would make it possible to examine the effects of scrambling item sequence in one unit of a program on (*a*) error rate during acquisition, (*b*) performance on a subsequent unit of the program, and (*c*) performance on an achievement test which measured retention and transfer.

### *Subjects*

The 36 subjects were all pupils in the last month of the second grade in a public school. These 13 boys and 23 girls were selected from among 81 children who were present in three classes on the day a screening test, Test S, was administered. The final sample was selected from among those who met the initial screening criteria by subsequently dropping 5 additional children. Three children were dropped because of poor performance on the first two units of the program. Subsequently, 1 child in each experimental group was dropped because of procedural errors during the experimental treatment. Thus, the final sample consisted of two groups of 18 subjects each.

Both the experimental and control groups had mean ages of 7.6 years with standard deviations of .6 years. The mean IQ (Otis tests) of the control group was 107.0 with a standard deviation of 11.3, while for the experimental group, the corresponding figures were 107.4 and 14.4.

### *Instruments*

The research instruments included two forms of a specially developed self-instructional program, teaching machines for presenting the programs, and two tests.

### *Programs*

The self-instructional program, informal geometry for the second grade,[2] had been developed through a process of trial and revision as suggested by Skinner (1958). A validation study had established that the program produced a statistically significant amount of learning, but was not highly effective in promoting mastery of the material.

The program consisted of 180 items, or steps, grouped in five units: Alpha, Dots, Lines, Angles, and Triangles. Alpha was developed to use while teaching children how to manipulate the teaching machines and did not contain any geometry items. The other four units comprised a linear sequence in which concepts and notation taught in one unit were used and built upon in subsequent units. The program had a mean of 6.7 words per item. Only 99 different words were used in the program and the type-token ratio for the program as a whole was .08. The 15 second graders who went through the program in the validation study, had a median error rate of 22.2% for the total program and took a median time of 59.9 minutes to go through the items on their first exposure to them.

The only portion of the experimental program which differed from the standard form was the unit on angles. The angles unit was selected for experimental treatment because of its difficulty and because of its position in the program. Previous results indicated that the angles unit was the most difficult. Moreover, the fact that the unit on angles came midway in the program made it possible to compare the treatment groups with respect to acquisition before and after the experimental treatment, as well as during it.

In the standard program the unit on

---

[2] A 19-page technical report, including a specimen copy of this program, has been deposited with the American Documentation Institute. Order Document No. 7515 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

angles consisted of 60 items which were presented on three strips, each containing 20 items (see Figure 1). In the experimental form of the program the order of these three blocks of 20 items was held constant, while the order of the 20 items within each block was randomly scrambled. The rank-order correlations between standard and random item sequences in the three blocks were −.17, −.28, and −.07, respectively.

*Program Presentation.* The program was duplicated on 8½ × 42 inch strips of paper by the ozalid process and these strips were presented by means of Foringer 2002 teaching machines. The items were spaced so that it was impossible to see portions of two items at once. The children wrote their composed answers on a standard answer strip.

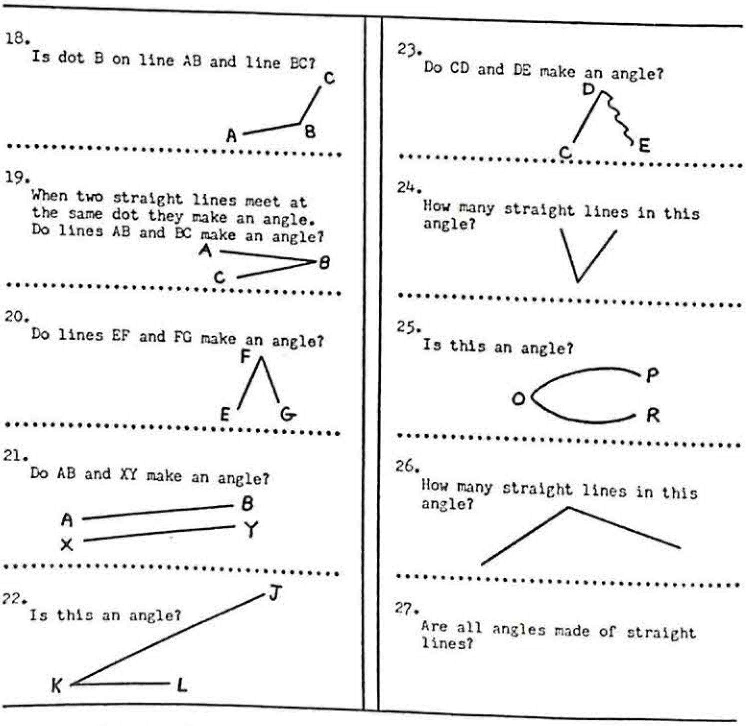
*Tests.* Two achievement tests had been developed for different purposes. Test S was designed for quick screening purposes. It was designed to screen out the subjects whose verbal comprehension was too poor for them to understand the program, and the subjects who were well acquainted with the material to be covered in the program. Test S consisted of 2 mimeographed pages containing 8 verbal comprehension items and 8 geometry items. Test 1 was developed to provide a means of assessing what a child learned by going through the program. The 52 items were grouped homogeneously on 7 mimeographed pages. Spearman-Brown corrected split-half reliabilities for Test S and Test 1 were .68 and .87, respectively.

## *Procedure*

The experiment began when the screening test was administered to the children in their classrooms. It ended 17 days later. During the experiment every subject participated for about 15 minutes every day, with only minor exceptions.

After the initial screening, all subsequent research sessions were carried out in a psychological laboratory in the basement of a nearby school. Most of the subjects had gone to first grade in this school and it was common for pupils to go to this school building on errands.

The subjects typically went from their classrooms to the laboratory in groups of four. Members of the experimental and control groups came in the laboratory simultaneously and in approximately equal numbers at all stages of the experiment. Two or three experimenters were in the laboratory at all times,



18.
Is dot B on line AB and line BC?

19.
When two straight lines meet at the same dot they make an angle.
Do lines AB and BC make an angle?

20.
Do lines EF and FG make an angle?

21.
Do AB and XY make an angle?

22.
Is this an angle?

23.
Do CD and DE make an angle?

24.
How many straight lines in this angle?

25.
Is this an angle?

26.
How many straight lines in this angle?

27.
Are all angles made of straight lines?

FIG. 1. Sample portion of the standard angles unit.

and they timed each child on each program segment with stop watches.

*Pretesting and Pretraining.* In their first session in the laboratory, the subjects were instructed, given the first 5 pages of Test 1, and taught to use the teaching machines while going through Unit Alpha of the self-instructional program. The next day pretraining began. During pretraining each subject went through the units on dots and on lines twice. After completing the second trial on lines, each subject was retested with 3 pages of Test 1 which covered the units on Dots, Lines, and Angles.

*Experimental Treatment.* The subjects were matched and assigned randomly to treatment groups on the basis of their scores on the angles items given during the first administration of Test 1. The preferred procedure of matching them on the basis of scores on both administrations of Test 1, IQ, and performance during pretraining had to be foregone due to time pressures.

There were no cues except for those provided in the program to indicate to the subjects that the procedure was different during the experimental treatment than previously. Throughout the experiment, the subjects were told that different children were working on different material.

During the experimental treatment each subject went through each segment of the angles program three times before moving on to the next segment.

*Posttraining and Posttesting.* After finishing the angles unit, all subjects were taken through the final unit of the program, Triangles, presented in standard form. All of the subjects finished the program on a Friday. On the following Monday, they all were given Test 1 and Test S.

## *Data to be Examined*

In order to determine the importance of the item sequence variable, several comparisons were necessary. Comparisons of the experimental and control groups with respect to pretest and pretraining performance would determine whether they were performing similarly before the introduction of the experimental treatment. An analysis of speed and error data during the experimental treatment would then be required to determine whether the treatment was related to acquisition performance. Performance on the Triangles unit of the program, scores on posttests, and comparisons of pretests and posttests would indicate whether retention and transfer criteria of learning showed effects of the experimental treatment. Finally, correlations between IQ, pretest scores, program scores, and posttest scores for each group would reveal how response measures interrelated, and whether such correlations would clarify group comparisons.

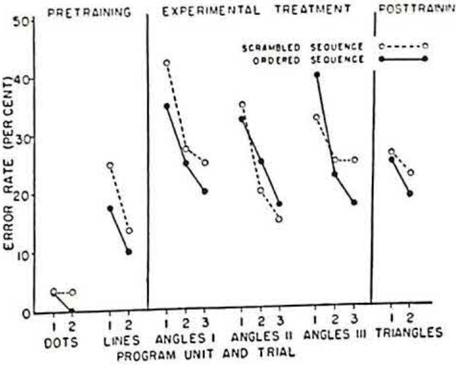


FIG. 2. Median error rates for each trial on each program unit.

## RESULTS

The data indicated that program performance through the experiment was similar for the experimental and control groups. Table 1 summarizes time and error score for pretreatment, treatment, and posttreatment performance on the program. Mann-Whitney $U$ tests indicated that none of the differences in Table 1 reached statistical significance. Figures 2 and 3 facilitate a more detailed examination of program perform-

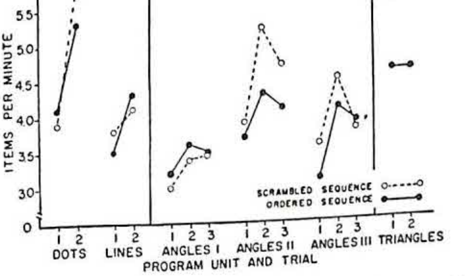


FIG. 3. Median speed for each trial on each program unit.

TABLE 1

PROGRAM PERFORMANCE

| Stage of training | Median errors | | | Median working time (minutes) | | |
|---|---|---|---|---|---|---|
| | Control group | Experimental group | $T$[a] | Control group | Experimental group | $T$[a] |
| Pretreatment | 13.5 | 18.5 | 313.0 | 32.2 | 32.8 | 327.5 |
| Treatment | 47.5 | 50.5 | 321.0 | 46.9 | 49.5 | 332.0 |
| Posttreatment | 16.5 | 19.0 | 308.0 | 17.9 | 14.8 | 282.0 |

[a] Smaller rank total. A rank total of 333.0 would indicate identical sums of ranks for the two groups. A rank total of 271.0 would provide a $p$ value of .05.

ance. These figures bear out the gross comparisons in Table 1.

Table 2 makes it possible to see the effects of the experimental treatment on both specific and general measures of knowledge of geometry. Mann-Whitney $U$ tests indicated that none of the comparisons approached statistical significance. Additional comparisons of such facets of test performance as transfer items, nontransfer items, and individual pages of the posttest also failed to indicate any differences in performance between the two groups.

Thus, group comparisons of acquisition, retention, and transfer data all indicated that the experimental treatment did not have statistically significant effects.

TABLE 2

TEST PERFORMANCE

| Test measure | Control median number correct | Experimental median number correct | Smaller rank total[a] |
|---|---|---|---|
| Angles Pretest 1 | 2.5 | 2.0 | 332.5 |
| Angles Pretest 2 | 3.0 | 3.0 | 308.5 |
| Angles posttest | 8.0 | 7.0 | 310.5 |
| Angles gain | 4.2 | 5.0 | 326.5 |
| Total Pretest 1 | 31.0 | 29.2 | 306.0 |
| Total posttest | 50.2 | 47.0 | 316.0 |
| Total gain | 18.0 | 18.2 | 333.0 |

[a] A rank total of 333.0 would indicate identical sums of ranks for the two groups. A rank total of 271.0 would have a $p$ value of .05.

Correlations were computed for several response measures to clarify the meaning of the group comparisons. In general it was found that measures of intelligence and quality of performance in geometry tended to be correlated consistently, while measures of speed of performance and gain from pretests to posttests were less consistent. Table 3 summarizes the correlations for intelligence and three gross measures of geometry knowledge.

## DISCUSSION

While the meaning of the results is, of course, subject to varied interpretation, the data certainly provide no

TABLE 3

SELECTED CORRELATIONS
($N = 18$ for each group)

| Variables[a] | Group | |
|---|---|---|
| | Control | Experimental |
| IQ and pretest | .24 | .50* |
| IQ and program | .68** | .58** |
| IQ and posttest | .45* | .56** |
| Pretest and program | .52* | .48* |
| Pretest and posttest | .52* | .65** |
| Program and posttest | .80*** | .51* |

[a] Pretest, posttest, and program scores refer to number of items answered correctly.
* $p \leqq .05$, one-tailed test.
** $p \leqq .01$, one-tailed test.
*** $p \leqq .0005$, one-tailed test.

support for the contention that item sequence is an important variable in self-instructional programs. There was no evidence that item scrambling impaired learning. Moreover, the most reliable of the many statistically insignificant comparisons was in favor of the experimental group.

The correlations between measures of individual differences at the beginning of the experiment with indices of learning suggest that individual differences might have obscured treatment effects. Analysis of the present data also indicated a slight tendency for bright children in the control group to do better than bright children in the experimental group, while the opposite trend appeared for the duller children. Thus, it seems clear that the possibility of interaction between initial ability and treatment effects deserves explicit investigation. However, it should be noted that the subjects in the present sample probably were no more variable with respect to intelligence than the children in an average American classroom. Thus, individual differences probably would be as important in many classroom settings as in the present experiment.

Probably the most important limitation of the present study was the failure of the program to teach the material presented thoroughly to most of the subjects, as indicated in posttest performance. It seems likely that a more effective program would have been made less effective when scrambled.

While the present findings failed to support the assumption that item sequence is important, it seems neither appropriate, nor even tempting to abandon the hypothesis that the order of presentation matters under some conditions. Instead, it seems useful to try and conceptualize the problem more adequately and operationally.

It is intriguing to think of item scrambling as a means of dissecting a program so as to learn more about its sequential structure. When we hold the sequence of units, or blocks of items, constant and vary item order within blocks, as in the present experiment, we are in a position to learn about the "fine structure" of the program. When we reorder blocks of items and hold order constant within these blocks, we can explore the "gross anatomy" of the program. A program consisting of totally unrelated items presumably would be found to have "no sequential structure," as indicated by no differences in indices of learning, regardless of what we did to the order of items. In contrast, a program with "completely linear structure" would be less effective if we tampered with its sequential structure in any manner, or to any degree.

It seems likely that such dissection would reveal that most existing programs are neither completely linear, nor completely lacking in sequential structure. Moreover, such dissection might lead to a useful means of describing and comparing the sequential structures of programs. Finally, we may only know how much emphasis to put upon sequence in programing when we can empirically characterize the sequential structure of self-instructional programs and have sound empirical generalizations about the relationships between sequential structure, content, and dependent variables.

## REFERENCES

GAVURIN, E. I., & DONAHUE, VIRGINIA M. Logical sequence and random sequence. *Automated Teach. Bull.*, 1961, **1**, 3–9.

HOLLAND, J. G., & SKINNER, B. F. *The analysis of behavior.* New York: McGraw-Hill, 1961.

SKINNER, B. F. Teaching machines. *Science*, 1958, **128**, 969–977.

(Received June 14, 1962)

*Journal of Educational Psychology*
1963, Vol. 54, No. 3, 144–148

# DECAY OF IMMEDIATE MEMORY FOR VISUALLY PRESENTED DIGITS AMONG NONREADERS AND READERS[1]

LINDA F. ALWITT

*Study Center, Pediatrics, Albany Medical College*

The purpose of this study was to examine 1 organism variable, rate of decay of memory traces, as a contributing factor in the low digit spans observed among cases of reading disability. 19 CA-matched reading disability and control cases were given instructions before and after a visual presentation as to which line of a matrix of 8 digits to report. The lack of difference between the groups of Ss suggests that immediate memory traces of reading disability cases do not decay at a faster rate than do those of normal readers. An incidental finding was that normal readers show a greater rate of increase in digit span with CA than do reading disability cases. Implications of both findings were discussed.

Children with reading disability have often been reported to exhibit low auditory memory spans (e.g., Rose, 1958; Saunders, 1931) and low visual memory spans (e.g., Rizzo, 1939; Rudisill, 1956). They also show low visuotemporal spans as measured by the visuokinesthetic Knox Cube Test (Robinson, 1946, p. 153) and by tests which use successive rather than simultaneous presentations of visual stimuli (e.g., Hawkins, 1897; Rizzo, 1939). Since these children have low spans in visual, auditory, and kinesthetic modalities, it is likely that their deficit in digit span lies in intervening processes, including immediate memory, rather than in sensory input processes. Inasmuch as digit span is associated with reading ability, it is important to ascertain the causes of the low digit spans observed in cases of reading disability.

The typical measure of memory span consists of the presentation of items to be recalled immediately. Among the organism variables which limit the memory span are: (*a*) the amount of material which the subject is capable of bringing to the processes of immediate memory, (*b*) the size of the subject's capacity for storing immediate memory traces, (*c*) the rate of decay of immediate memory traces, and (*d*) the rate of reporting trace material. The current study is primarily concerned with examining one of these organism variables, rate of decay of memory traces, among reading disability cases and normal readers.

Brown (1958) and Sperling (1960) have presented evidence that immediate memory traces decay with time among presumably normal readers. The lowered digit span observed among reading disability cases can be due to a faster rate of decay of immediate memory traces, compared with normal readers.

The method chosen to study trace decay was adapted from Sperling (1960), who sampled reports of stimuli seen in brief visual presentations. Subjects were auditorially cued as to which line of a visual matrix of letters and/or

---

[1] This study was supported in part by Public Health Service Project Grant OM 225 and by a research grant from the Association for the Aid of Crippled Children. Appreciation is expressed to N. D. Bryant for his helpful comments on a draft of this paper.

digits to report, and samples of performance were obtained. Sperling found that immediate memory for the visual presentations decreased as a function of a delay in cue instructions as to which part of the stimulus matrix to report. This result may be interpreted as a reflection of the decay of immediate memory traces with time.

## METHOD

### *Subjects*

Nineteen children with reading disabilities, who were entered in the diagnostic program of the Albany Study Center for Learning Disabilities, were tested. The proportion of boys to girls in this sample (17 boys and 2 girls) is in accord with the proportion found among serious reading disability cases (Vernon, 1957). The inclusion of girls in this sample should not affect the results since no sex differences have been found for children with regard to memory span (Blankenship, 1938). The mean Wechsler Intelligence Scale for Children, Full Scale IQ for this group is 98.4 ($SD = 9.0$). The average reading ability of this group is 1.5 years below their grade placements on the basis of the Gilmore Oral Reading Test.

Because memory span has been shown to increase with age within this age range (Hawkins, 1897; Jacobs, 1887), reading control cases were individually matched with reading disability cases on the basis of chronological age. The control cases were chosen from a population of 60 public school boys enrolled in Grades 1 through 6. As tested by school personnel, these boys were of average or above average intelligence as measured by the Otis Alpha or Beta, and were at or above their grade placements in reading ability. Table 1 shows the average chronological ages and grade placements for the two groups.

TABLE 1

CHRONOLOGICAL AGE AND GRADE PLACEMENT OF THE READING DISABILITY AND CONTROL GROUPS

| Sample characteristics | Reading group | |
|---|---|---|
| | Disability | Control |
| Chronological age (in months) | | |
| *M* | 127.2 | 127.5 |
| *SD* | 15.5 | 15.7 |
| Grade placement | | |
| *M* | 4.0 | 4.8 |
| *SD* | 1.1 | 1.2 |

### *Material*

Eight digits were typewritten in 3/16 inch Ampli type on white cards. The digits were arranged in two rows of four digits each. Eight of the Arabic numerals 1 through 9 appeared once in each matrix, and an equal number of digits appeared in all of the 24 cards used in the study. A Lafayette Tachitron was used to present the cards at a .1-second exposure time. The area of the card which was visible in the viewing aperture was 5 3/16 inches wide × 3 5/16 inches high, and was viewed from a distance of about 8½ inches.

### *Procedure*

Sperling's (1960) results indicated that there was little difference between the amount of material available when cuing instructions were given .05 second before the stimulus presentation, and at the same time as the stimulus presentation. Based on this result, and in order to simplify the experimental technique, the two partial reporting conditions of the current study employed only precuing and postcuing instruction conditions. The precuing instruction condition served to *estimate* presentation of the cuing instruction and the stimulus simultaneously. Therefore effects of attentional set, arising from instructions to pay attention to four stimuli in the precuing condition and eight stimuli in the postcuing condition, were not considered in the analysis of the experimental results.

In the current study, before the digits were presented, the subject was shown a matrix of eight boxes corresponding to the placement of the digits. He was told that eight digits would be presented for a very brief time. After he saw the digits, his task was to point to each box on the response matrix and orally report the digit he had seen there. Guessing was encouraged.

Condition I consisted of the presentation of four cards with instructions to report all eight digits in any order. The presentation of each stimulus was preceded by 2 seconds with the oral signal "ready."

Condition II consisted of the presentation of eight cards with instructions to report the digits of only one of the rows. About 1.5 seconds after the ready signal, the experimenter orally cued the subject as to which row to report by saying "top" or "bottom," and then immediately presented the stimuli. Half of the cards were preceded by each instruction, and the instructions were varied in a random fashion so that guessing the instruction was minimized.

Condition III consisted of the presentation of eight cards with instructions to report the digits of only one of the rows. However, the cuing instruction was given about .5 second after the presentation of each stimulus. The stimulus was always preceded by the signal "ready."

Condition IV was the same as Condition I.

The four conditions were always presented in the same order. Since practice effects do not increase digit span within a testing session (Sperling, 1960), it was felt that a balanced order of presentation of Conditions II and III was not essential. However, Conditions I and IV were compared in order to detect any practice effects within the testing session.

Every card was scored for the number of items reproduced in their correct positions. For each condition of each subject, a mean digit span was computed. The mean digit span is defined as the proportion of correct letters within a condition multiplied by the number of digits requested. Thus the proportion is multiplied by eight in Conditions I and IV, and by four in Conditions II and III.

## RESULTS

In order to assess the combined effects of (*a*) practice in reporting digits and (*b*) familiarity with the experimental procedure, Conditions I and IV were compared. Bartlett's test for homogeneity of variance was not significant, so raw data were used. Table 2 shows the mean spans for the groups of subjects in Conditions I and IV. There was no significant difference in mean digit span between the beginning (Condition I) and the end (Condition IV) of the testing session ($F=1.14$, $df=1/36$). This finding suggests that the confounded effects of practice and familiarity were insignificant. The results of Conditions II and III are also assumed to be free from influence by these effects.

TABLE 2

MEAN DIGIT SPANS FOR EACH GROUP IN EACH CONDITION

| Condition | Reading group | |
|---|---|---|
| | Disability | Control |
| I | 1.37 | 2.61 |
| II | 1.72 | 2.54 |
| III | 1.24 | 1.80 |
| IV | 1.51 | 2.98 |

Decay of immediate memory traces is reflected in the comparison of pre- and postpresentation cuing instructions. Because of the delay of instructions as to which line of the stimulus matrix to report in the postcuing condition (III), it was expected that the mean digit span would be lower in this condition that in the prepresentation cuing condition (II). The extent of decay of immediate memory traces is reflected by the amount of difference between the two cuing conditions. The relationship between the prepresentation cuing instruction condition (II) and the postpresentation cuing condition (III) for the groups of subjects was compared by means of an analysis of variance. Since Bartlett's test for homogeneity of variance was not significant, raw data were used. Table 2 shows the mean digit spans for each group in Conditions II and III. The reading control group showed a higher mean digit span than the disability group in both cuing conditions ($F=7.84$, $df=1/18$, $p<.05$). Prepresentations cuing instructions yielded a higher mean digit span than did postpresentation cuing instructions ($F=34.75$, $df=1/37$, $p<.001$). However, the difference between the experimental conditions was not greater for the reading disability group than for the control group, contrary to expectations ($F=1.12$, $df=1/18$).

Of secondary interest in this study is the significant first order interaction between pairs of subjects and groups of subjects, which was observed in both analyses of variance (Condition I versus IV: $F=4.90$, $df=18/37$ $p < .001$; Condition II versus III: $F=5.70$, $df=18/37$, $p < .001$). Since subjects were individually paired on the basis of chronological age, this interaction is largely due to age. Figure 1 presents the relationship of mean digit span to chronological age separately for each group of subjects. Each group was subdivided into five age groupings, and lines were fitted to the two sets of data by the method of least squares. Reading disability subjects of age 110 months lagged about 8 months behind reading control subjects in digit span, but disability subjects of age 140 months lagged about 21 months behind the reading controls in span. The trends suggest that mean digit span increases with chronological age more rapidly for the control group than for the reading disability group.

## DISCUSSION

The results of this study confirm Sperling's report in that the number of items available in the prepresentation cuing conditions was greater than the number of items available in the postpresentation cuing instruction condition. This implies that the same processes were tapped in the current study as were tapped in Sperling's study. Furthermore, the lowered memory span so often observed among cases of reading disability was also observed in the current study.

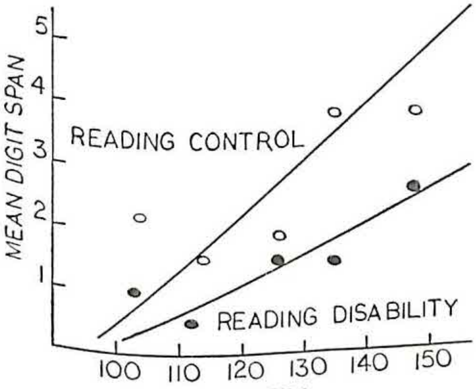


FIG. 1. Relation of mean digit span to chronological age in reading disability and control groups.

The lack of difference between the groups of subjects in the number of items available in precuing and postcuing instruction conditions suggests that the immediate memory traces of reading disability cases do not decay at a faster rate than do those of normal readers. Rather, the underlying mechanism of the lowered digit span among reading disability cases must be sought in the other organism variables, mentioned previously, which are related to immediate memory. For example, reading disability cases may not pay attention to the entire stimulus field, and thus all of the elements of the stimulus will not reach the immediate memory processes. Another possibility is that reading disability subjects, compared with normal readers, have a stricter limit on the amount of material which they are capable of holding in the immediate memory. A third possibility is that, since reading disability cases often display aphasoid characteristics (Bryant & Patterson, 1962), they also have difficulty in associating a given memory trace with the verbal response element which represents this trace, and with responding with this verbal element. During the longer time required by them to make the association and response, the immediate memory trace may decay below the threshold of emission. Thus trace decay may play a role in the lowered digit span of reading disability cases only because of the influence of response factors on it.

The increase of digit span with an increase in chronological age was expected on the basis of studies previously mentioned. However, the greater rate of increase observed among the normal readers compared with the reading disability cases was unforeseen. This difference in rate of increase in span between levels of reading ability can be interpreted in two ways. First, practice in reading may incidentally increase memory span by providing practice in perceiving and remembering large chunks of material. Since the effect is incidental, no gain in reading ability would be expected as a result of transfer from practice in tachistoscopic viewing of a series of digits. This interpretation of the trends implies that prospective reading disability cases have the same span abilities as do prospective good readers, before reading instruction is given. A second interpretation of the divergent age trends is that the low span of reading disability cases is inherently associated with the substructure which underlies their reading difficulty. This interpretation implies that low memory span contributes to reading disability in that the child is unable to remember a sufficient number of stimulus elements to utilize them for reading. The view of lowered digit span of poor readers as a function of deficit in attentional, immediate memory, or response processes is compatible with both interpretations of the divergent age trends.

## REFERENCES

BLANKENSHIP, A. B. Memory span: A review of the literature. *Psychol. Bull.*, 1938, **35**, 1–25.

BROWN, J. Some tests of the decay theory of immediate memory. *Quart. J. exp. Psychol.*, 1958, **10**, 12–21.

BRYANT, N. D., & PATTERSON, P. R. Reading disability: Part of a syndrome of neurological dysfunctioning. Paper read at International Reading Association, San Francisco, May 1962.

HAWKINS, C. J. Experiments on memory types. *Psychol. Rev.*, 1897, **4**, 289–294.

JACOBS, J. Experiments on "prehension." *Mind*, 1887, **12**, 75–79.

RIZZO, N. D. Studies in visual and auditory memory span with special reference to reading disability. *J. exp. Educ.*, 1939, **8**, 208–244.

ROBINSON, HELEN M. *Why pupils fail in reading.* Chicago: Univer. Chicago Press, 1946.

ROSE, FLORENCE C. The occurrence of short auditory memory span among school children referred for diagnosis of reading difficulties. *J. educ. Res.*, 1958, **51**, 459–464.

RUDISILL, MABEL. Flashed digit and phrase recognition of oral and concrete responses: A study of advanced and retarded readers in the third grade. *J. Psychol.*, 1956, **42**, 317–328.

SAUNDERS, MARY JANE. The short auditory span disability. *Childh. Educ.*, 1931, **8**, 59–65.

SPERLING, G. The information available in brief visual presentations. *Psychol. Monogr.*, 1960, **74**(11, Whole No. 498).

VERNON, M. D. *Backwardness in reading.* Cambridge: Universities Press, 1957.

(Received November 7, 1962)

*Journal of Educational Psychology*
1963, Vol. 54, No. 3, 149–153

# FACTORS RELATED TO THE SHIFT OF PROFESSIONAL ATTITUDES OF STUDENTS IN TEACHER EDUCATION

LEON T. OFCHUS AND WILLIAM J. GNAGEY

*Illinois State Normal University*

The 71 college sophomore women enrolled in a teacher education curriculum were given the MTAI before and after a 1-semester course in child growth. Using the California F Scale and a personal rating form, inquiry was made into the correlates of attitude shifts during the course. Although student perception of the instructor's competence, permissiveness, and attitudes seemed not to be related to attitude shift, student ratings of their own parents were significantly associated with their perception of the instructor. A negative correlation ($r = -.58$) was found between the California F Scale and the MTAI.

Evidence of the effects of teachers' attitudes upon their students has long been available. As early as 1934, Krall found that students' social attitudes tended to shift toward the position held by their teacher. Recently Della Piana and Gage (1955) found a positive relationship between a teacher's attitudes as measured by the Minnesota Teacher Attitude Inventory (MTAI) and the value systems of his pupils. More recently, Hoyt and Cook (1959) found a positive relationship between the teacher's MTAI score and her pupils' ratings of the whole school experience.

The present study was designed to explore some relationships between the student-class-course-instructor complex and changes in student attitudes toward teaching children as measured by the MTAI.

## Some Theoretical Considerations

It was postulated that a college instructor may influence the attitudes of his students without direct, forthright, and planned efforts to do so. The designation of a particular text, the choice of lecture and discussion topics, the selection of films: through all these the instructor reveals repeatedly his underlying attitudes toward students and teaching. Even more potent could be the fact that a majority of the attitudes measured by the MTAI are expressed daily by the character of the instructor's reaction to and interaction with his own students.

If a student senses a sizable breach between his own set of attitudes and his instructor's manifest point of view, a state of dissonance may be produced in the psychic structure of that student (Festinger, 1957). After all, the college instructor is a person who can commandeer nearly all the sources of social power cited by French (1960). He is an expert; he possesses the power to reward and punish in a significant way; he has the legitimate right of the teacher to exert influence; and, if his MTAI score is high and valid, he may be an extremely attractive individual in addition.

There are several ways that a student may attempt to reduce this dissonance: (*a*) he may leave the course, (*b*) he may shift his attitudes toward those of his instructor, (*c*) he may perceive the instructor as an inferior teacher un-worthy of emulation, (*d*) he may reject teaching itself as a long term career for him. As a consequence of this theorizing, we predicted the following relationships:

1. Students who make large instructorward attitude shifts will perceive the instructor as more permissive, more capable, and nearer to their own attitudinal position than will students who make minimal or opposite shifts.

However, a student's perception of his instructor may be distorted by some antecedent factors that could obliterate many of the pure effects of the instructor's influence. One major dynamic may be the sheer displacement of feelings for a parent on to the instructor. As a result of these conjectures we predicted the following:

2. Students who feel that their parents are either highly concerned about them or relatively uncritical of their behavior will perceive the instructor as significantly more permissive than will students who feel that their parents are lacking in concern for them or were highly critical of their behavior.

Since the writers of the MTAI have used the characteristics of the "authoritarian personality" (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950) as a sort of operational definition of the "poor teacher" (Cook, Leeds, & Callis, 1951), we predicted that:

3. There will be a negative relationship between authoritarianism as measured by the California F Scale and positive attitudes toward children as measured by the MTAI.

Since rigidity of position has been strongly associated with authoritarianism (Adorno et al., 1950) we predicted that:

4. Students who have highly authoritarian personalities will make fewer high instructorward shifts than students who are low in the authoritarian factor.

Logically, the picture of the "good teacher" inadvertently presented by an instructor with "high MTAI attitudes" should produce more dissonance in students whose attitudes score at the lower end of the MTAI that in those who score high. We subsequently predicted that:

5. Students who have low final MTAI scores will less often plan to make a career of teaching than will students who have high final MTAI scores.

## RESEARCH DESIGN

### *Subjects*

Seventy-one students at Illinois State Normal University who were enrolled in a required course in two sections of child growth and development during the second semester of 1961–62 served as subjects. With few exceptions, the subjects were college sophomore women enrolled in either elementary education or special education curriculums.

Both sections were taught by the same instructor who followed essentially the same instructional procedures in each. The MTAI score of the instructor was 99, considerably above the average for his education and experience.

### *Measurement of Variables*

At the beginning of the semester, the MTAI and the California F Scale were administered to all subjects.

Toward the end of that semester, the MTAI was readministered along with a rating scale on which each student indicated, among other things, her perception of (*a*) the instructor's permissiveness of disagreement with his opinion, (*b*) the instructor's ability to explain material meaningfully, (*c*) the congruence of her own and the instructor's attitudes about children, (*d*) the criticalness of her own father and mother, and (*e*) the level of concern shown her by her own father and mother.

An additional question asked each student to indicate whether or not she thought she would be teaching 10 years from now.

### *Scoring*

The rating scales were scored by measuring the number of millimeters between the ex-

TABLE 1

CHI SQUARE COMPARISON OF INSTRUCTOR RATINGS BY STUDENTS WHO MADE HIGH INSTRUCTORWARD MTAI SHIFTS WITH THOSE BY STUDENTS WHO MADE MINIMAL OR OPPOSITE SHIFTS

| | Instructor permissiveness | | | | Instructor competence | | | | Congruence with instructor attitude | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Shift | Above median | Below median | $\chi^2$ | $p$ | Above median | Below median | $\chi^2$ | $p$ | Above median | Below median | $\chi^2$ | $p$ |
| High | 19 | 15 | | | 20 | 14 | | | 19 | 15 | | |
| Low | 12 | 19 | 1.92 | .08 | 13 | 19 | .30 | .25 | 13 | 19 | 1.54 | .13 |

treme right of the scale and the farthest edge of the subject's vertical mark. Attitude shifts were measured by noting the magnitude and direction of the pre-post differences in a student's MTAI score.

The teaching career question was scored either yes or no, while both the California F Scale (used to measure authoritarianism) and the MTAI were scored according to the standard procedure indicated in each manual.

The chi square median test was used in nearly every analysis.

## RESULTS

### *Hypothesis 1*

As can readily be seen in Table 1, no significant differences were found between the high shifters' and low shifters' ratings of instructor permissiveness, instructor competence, or in the student-instructor congruence of attitudes toward children. We were unable to reject the null hypothesis in any of these cases.

### *Hypothesis 2*

When students who rated the intructor as highly permissive of disagreement were compared with students who rated him low in permissiveness, several relationships emerged suggesting antecedent forces.

1. Table 3 shows that students who rated their mothers as relatively uncritical of them perceived their instructor to be significantly more permissive than did students who rated their mothers as relatively supercritical of them ($\chi^2 = 5.07$, $df = 1$, $p < .025$). Table 2 indicates that the same was true for a similar rating of their fathers ($\chi^2 = 4.14$, $df = 1$, $p < .025$).

Parenthetically, a contingency coefficient of correlation was computed between the students' ratings of the criticalness of their mothers and their fathers. A sizable positive relationship

TABLE 2

CHI SQUARE COMPARISONS OF FATHER RATINGS BY STUDENTS WHO RATED THE INSTRUCTOR HIGH IN PERMISSIVENESS WITH THOSE BY STUDENTS WHO RATED HIM LOW IN PERMISSIVENESS

| | Criticalness of father | | | | Concern of father | | | |
|---|---|---|---|---|---|---|---|---|
| Instructor permissiveness | Above median | Below median | $\chi^2$ | $p$ | Above median | Below median | $\chi^2$ | $p$ |
| High | 13 | 20 | | | 16 | 16 | | |
| Low | 23 | 13 | 4.14 | .025 | 17 | 17 | 0 | 100 |

TABLE 3

CHI SQUARE COMPARISONS OF MOTHER RATINGS BY STUDENTS WHO RATED THE INSTRUCTORS HIGH IN PERMISSIVENESS WITH THOSE BY STUDENTS WHO RATED HIM LOW IN PERMISSIVENESS

| | Criticalness of mother | | | | Concern of mother | | | |
|---|---|---|---|---|---|---|---|---|
| Instructor permissiveness | Above median | Below median | $\chi^2$ | $p$ | Above median | Below median | $\chi^2$ | $p$ |
| High | 11 | 20 | | | 23 | 11 | | |
| Low | 21 | 12 | 5.07 | .025 | 12 | 24 | 8.24 | .005 |

was found ($C=.39$). This relationship was significant at better than the .001 level of confidence ($\chi^2=12.82$, $df=1$).

2. Table 3 reveals that students who rated their mothers as being highly concerned about their well-being perceived the instructor to be significantly more permissive than students who rated their mothers as relatively low in concern for them ($\chi^2=8.24$, $df=1$, $p < .005$). However, Table 2 shows that this relationship did not hold for ratings of their fathers.

### *Hypothesis 3*

When a product-moment correlation coefficient was computed between the scores on the California F Scale and the MTAI scores for the 71 students, a sizable negative relationship was found ($r=-.58$). This relationship was significant at better than the .01 level of confidence ($t=6.0$, $df=70$).

### *Hypothesis 4*

No significant difference in authoritarianism (as measured by the California F Scale) was found between high shifters and low shifters.

### *Hypothesis 5*

Table 4 shows that students who made low final MTAI scores planned significantly less often to be teaching in 10 years than did students who made high final MTAI scores ($\chi^2=3.16$, $df=1$, $p < .04$). When the initial MTAI scores of the same two groups were compared, no significant difference was apparent ($\chi^2=.09$, $df=1$, $p > .70$).

## CONCLUSIONS AND DISCUSSION

Under the conditions of this study, neither the students' authoritarianism, nor their perception of their instructor's competence, permissiveness, and

TABLE 4

CHI SQUARE COMPARISON OF MTAI SCORES OF STUDENTS WHO PLAN TO BE TEACHING IN 10 YEARS AND STUDENTS WHO DO NOT

| | Initial MTAI scores | | | | Final MTAI scores | | | |
|---|---|---|---|---|---|---|---|---|
| Plan to be teaching in 10 years? | Above median | Below median | $\chi^2$ | $p$ | Above median | Below median | $\chi^2$ | $p$ |
| Yes | 21 | 21 | | | 25 | 16 | | |
| No | 13 | 15 | .09 | $> .70$ | 11 | 18 | 3.61 | $< .04$ |

agreement with their own attitudes toward children were associated with a change in the students' professional attitudes as measured by the MTAI. It is possible that students encountered influences in other courses taken concurrently with child development that counteracted any tendency to shift toward the child centered attitudes of the experimental instructor. It is also possible that their perception of the instructor was distorted by some antecedent variables.

A major finding of this study was relationships suggesting a tendency for students to transfer on to the instructor their feelings about a mother-father authority figure. It may be that perception of an instructor was so strongly a function of this phenomenon that his manifest classroom behavior was irrelevant to attitude change.

As was predicted, it appears that students with high authoritarian tendencies seem to have very few warm, accepting attitudes toward children.

As predicted, students who came up with high final MTAI scores were the ones who planned to be teaching in 10 years. The dissonance caused by students' low initial scores and the instructor's child-centered philosophy may have been reduced by some students by rejecting teaching instead of by shifting toward the instructor's attitudinal position.

## REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

COOK, W. W., LEEDS, H., & CALLIS, R. *Minnesota Teacher Attitude Inventory*. New York: Psychological Corporation, 1951.

DELLA PIANA, G. M., & GAGE, N. J. Pupils' values and the validity of the Minnesota Teacher Attitude Inventory. *J. educ. Psychol.*, 1955, **46**, 167–178.

FESTINGER, L. *A theory of cognitive dissonance*. Evanston, Ill.: Roe, Peterson, 1957.

FRENCH, J. R. P., & RAVEN, B. The bases of social power. In D. Cartwright & A. Zander (Eds.), *Group dynamics*. Evanston, Ill.: Roe, Peterson, 1960.

HOYT, C. J., & COOK, W. W. The predictive value of the MTAI based on pupil attitude toward school. *J. teacher Educ.*, 1959, **10**, 42–44.

KRALL, A. The teacher's influence upon the social attitudes of boys in the twelfth grade. *J. educ. Psychol.*, 1934, **25**, 274–280.

(Received November 7, 1962)

*Journal of Educational Psychology*
1963, Vol. 54, No. 3, 154–159

# STRATEGY SELECTION AND NEGATIVE INSTANCES IN CONCEPT LEARNING[1]

LOY S. BRALEY

*University of California, Santa Barbara*

Exclusion solutions were investigated by requiring Ss to utilize information from negative (nonreinforced) instances in a concept learning task. Of 70 exclusion condition Ss only 3 adopted an exclusion strategy. Task performance was unrelated to conditions of reinforcement, instruction, and type of concept. Failure to use negative information was related to fundamental information handling strategies such that limited memory storage induces trial-and-error search based on positive identification of reinforced instances with complex stimulus arrays. A reduction of the cue space should favor use of strategies based on memory storage of what is not correct.

Recent and current interest in computer simulation of problem solving and concept learning, as well as the burgeoning area of programed instruction, have given increased impetus to a more systematic analysis of problem solving processes. The fidelity of any simulation rests, of course, on knowledge of how human subjects do in fact go about solving problems, although contributions to understanding occur reciprocally. The present investigation concerns one of the boundary problems in the area, viz., discounting storage and speed advantages of computers, are there problem solving strategies or information processing procedures that could readily be built into machines but which have only a tenuous analog in human strategy selection? One area where this may well be the case is that of the processing of negative information, where negative information refers to information as to what is *not* correct with regard to a given problem situation. Mention has been made, and some evidence presented (Bruner, Goodnow, & Austin, 1956; Donaldson, 1959; Hovland & Weiss, 1953) to suggest that human subjects are apparently either unwilling or unable to effectively utilize negative information under conditions where its informational value, as far as problem solving is concerned, is equivalent to positive information. If this is so, there are very practical implications for education as well as for the understanding of the learning process and its simulation.

The problem solving behavior that was the focus of this investigation was that of concept learning where negative information was represented by negative instances of the concept to be learned. As far back as 1933 Smoke had concluded that negative instances contributed little to concept learning. More recently Hovland and Weiss (1953), Friebergs and Tulving (1961), and Cahill and Hovland (1960) were able to demonstrate that concept learning, though difficult, was possible with negative instances, but apparently only when the subjects are provided with a complete specification of the population of cues (number and value of relevant dimensions) prior to solution attempts. The question of the ability or willing-

---

[1] This research was supported by Grant MH-06608 from the National Institute of Mental Health, United States Public Health Service.

ness of the subjects to utilize negative information under the usual concept learning conditions has not been adequately explored.

There are two related, but separable, aspects of the concept learning process that require mention. The first of these is the discrimination of relevant cues from some larger space of cues, upon which final problem solution rests. Secondly, the discriminations that occur depend on the strategy employed to select and test cue information. In many cases the use of negative information actually refers to the subject's use of an exclusion strategy rather than the use of negative cues per se. Exclusion strategies are familiar in mathematics where proof sometimes must proceed by the exclusion of other members of a given set that are not true, rather than by positive identification of what is true. Trouble shooting procedures furnish another example of where attempts at positive identification of the locus of malfunctioning is frequently far less efficient than proceeding by the exclusion of intact subsystems until the malfunctioning subsystem is identified by exclusion. One of the principal distinctions between strategies based on positive identification and those based on exclusion is that the latter require much greater use of memory processes while the former are more in the nature of trial and error.

A sequence of three experiments was conducted to explore the acquisition and utilization of negative information as a function of certain types of concepts, conditions of instruction, and conditions of reinforcement. The procedure in all studies entailed simultaneous presentation of a positive and negative instance on each trial. Conventionally in such tasks, each positive is an exemplar of the concept to be learned, while the negative instance has one or more critical dimensions of the concept absent with no consistent defining attribute except the absence of the correct concept. This need not be the case however, since it is a simple matter to employ two concepts with reinforcement defining the one to be learned. In this case the nonreinforced concept would then be represented by the "negative" instances. The subject could meet criterion either by learning the reinforced concept or learning the nonreinforced concept and choosing the opposite instance on each trial. In addition, the handling of negative information when only a single concept is present can be investigated by a procedure which reinforces as correct the subject's choice of those instances devoid of any consistent defining property. Both of these are referred to as exclusion solutions (or strategies)—that is, where the subject meets criterion by learning a concept represented by the negative instances and then chooses the opposite instance on each trial. To be consistent with current terminology (vide Kendler, 1961) exclusion solutions can be classified according to the type of concept employed in the negative instances. The conventional types would then be: conjunctive-exclusion, relational-exclusion, or disjunctive-exclusion solutions. In general, conjunctive concepts are those where each of the positive instances has the defining property or properties, while in disjunctive concepts the positive instances have one *or* the other of the defining properties, so at least two instances are necessary to define the concept. Relational concepts do not have any specific stimuli in common but rather the attributes on the positive instances have a common set of relationships.

## EXPERIMENT I

The first experiment was designed to investigate use of negative information in a task involving simultaneous presentation of an exemplar of each of two concepts on each trial. The subject could therefore meet criterion either by learning the reinforced concept or by learning the concept represented by the negative instances and then choosing the opposite instance. The information available on any trial is equivalent for either procedure. However, where the reinforced concept is the more difficult, greater efficiency could be achieved by learning the easier concept represented in the negative instances and then, by exclusion, choosing the one that is reinforced as correct; a very obvious procedure for a machine. To evaluate the possible contribution of set-to-solution attempts, instructions were manipulated to induce either a concept attainment or a general problem solving orientation.

### *Subjects*

The subjects were 40 high school students participating in a university level program of summer instruction, 22 of whom were females. All the subjects were volunteers and were randomly assigned to the four experimental conditions.

### *Design and Procedure*

The two conditions of instructions and the two types of concepts formed a 2 × 2 factorial design with 10 subjects per cell. On each trial all subjects were exposed to the simultaneous presentation of two 4 × 6 cards presented in adjacent windows of a large black plywood screen. A subject indicated his choice by depressing one of two keys affixed to the table immediately below the exposure windows. Upon a correct choice, reinforcement was delivered immediately by a light located at the upper rim of the exposure windows. Twelve successive correct responses were necessary to meet criterion.

Each of the two cards presented on any one trial contained two small geometric figures inside a larger geometric figure. The three figures on any card were combinations of circles, squares, and triangles, each of which could be red, blue, green, or yellow. One of the cards on each trial was an exemplar of a relational concept, and the other that of a disjunctive concept. Defining properties of the former were two similarly colored geometric figures within a differently colored larger geometric figure. The disjunctive concept was defined by two differently colored geometric figures *or* two similarly colored figures within a larger figure of the same color. On the basis of choice behavior alone it was not possible to determine if an exclusion solution had been employed and so an interrogation procedure was developed and administered after a subject had completed the task.

### *Instructions*

In the critical segment of the instructions half of the subjects were instructed to:

> discover what it is that makes a card correct, that is, find those characteristics of the cards that you need to pay attention to so as to be able to select the correct card on each trial [concept attainment instructions].

To avoid indicating that there were specific correct cards, the problem solving oriented instructions were:

> make a choice each time the two cards appear and the light [reinforcing stimulus] will tell you whether or not you are correct. In this way you will discover what it is that you need to pay attention to so as to reach a point where you can turn on the light on each trial.

### *Results*

Table I presents the means and standard deviations for the four groups

TABLE 1

TRIALS TO CRITERION AND STANDARD DEVIATIONS FOR RELATIONAL AND DISJUNCTIVE CONCEPT LEARNING IN EXPERIMENT I

| Instructions | Relational *M* | Relational *SD* | Disjunctive *M* | Disjunctive *SD* |
|---|---|---|---|---|
| Concept attainment | 14.9 | 10.5 | 30.3 | 16.1 |
| Problem solving | 17.5 | 12.1 | 30.6 | 17.3 |

showing clearly that the relational concept is easier to learn. Analysis of variance ($F = 5.37$, $df = 1/36$) on square root transformed trials to criterion yielded significance only for the type of concept. Analysis of the protocols indicated that despite the greater ease of attainment of the relational concept not a single subject made use of a relational-exclusion solution to meet criterion when reinforced for choice of the disjunctive concept. This failure to economize in information handling suggests that the subjects proceeded by positive identification of what was correct so that attention was directed largely to those cards that were reinforced and consequently the information available in the negative instances was either ignored or overlooked.

## EXPERIMENT II

This experiment was designed to increase assurance that the absence of exclusion solutions was not a shortcoming of the subject's report and to test more directly the ability of the subjects to make use of negative information in the form of exclusion strategies when reinforcement was delivered for the choice of instances that had no consistent defining property and a concept was present only in the negative (nonreinforced) instances.

### *Subjects*

Subjects for this experiment were drawn from the same population as in Experiment I. Sixteen males and 24 females were evenly distributed among the four conditions.

### *Procedure*

Major details of procedure and design were similar to Experiment I, including manipulation of instructions. However, in this experiment half of the subjects were reinforced on a conjunctive concept and the remaining half on those instances devoid of any consistent defining attribute, such that an exclusion procedure was the only possible mode of solution. The conjunctive concept was that of a small blue triangle. The dimensions of the instances were three values of figure (triangle, square, and diamond), four of color (red, blue, green, and yellow), and two of size (large, small).

### *Results*

Means for learning the conjunctive concept were 37.9 and 33.3 (*SD*s= 18.2., 17.8, respectively), with the problem solving instructional condition somewhat, but not significantly, easier. For those subjects where criterion could only be achieved through use of a conjunctive-exclusion solution, not a single subject was able to do so. However, in view of the possibly low termination of trials (60 trials) an additional 10 subjects were run for 140 trials on the exclusion procedure, 5 under each instructional condition. Only 1 of these 10 reached criterion and then could not verbalize his basis of choice. With only 1 subject out of 30 able to reach criterion, statistical analysis of these data appeared unnecessary.

Whereas the first experiment indicated failure to utilize negative information efficiently, this one demonstrated that problem solving occurs very infrequently when positively reinforced instances are devoid of any consistent defining attribute and an exclusion solution is the only one possible. There is, however, a possible source of contamination in Experiment II, depending on whether reinforcement operates symmetrically or asymmetrically. If the reinforcing stimulus differentially controls the attention of the subject to positive or negative instances, then the failure of exclusion strategies to emerge in Experiment II may be attributed to conditions of reinforcement rather than to a more general inability to utilize information from negative instances. Experiment III was designed to investigate this possibility.

## EXPERIMENT III

### *Procedure and Subjects*

This experiment was a partial replication of the second. The instructional variable was eliminated in light of the insignificant relationship to differential performance. Half of the subjects was reinforced on the conjunctive concept and the remaining half on those instances devoid of any consistent defining attribute, as in Experiment II. However, for half of each of these two groups reinforcement was delivered for a choice of the correct instance. For the other half of both groups the occurrence of the reinforcing stimulus signaled to the subject that he had made an incorrect choice and no light signaled that the correct choice had been made.

The 2 × 2 design again had four cells with 10 subjects per cell. The 40 subjects were volunteer undergraduates, 33 of whom were female. Stimulus materials were the same as in Experiment II and all subjects were given the problem solving instructions. The criterion remained the same for learning, but training was discontinued at 90 trials, instead of 60, if the subject had not met criterion.

### *Results*

Learning of the conjunctive concept proceeded much as in Experiment II for the problem solving instructional condition (*M*s = 31.0, 27.3; *SD*s = 17.2, 16.9). Of the 20 exclusion condition subjects only two met criterion—in 70 and 77 trials. Both of these subjects were in the condition where the reinforcing stimulus was delivered for a correct choice. These results constitute a replication of Experiment II, and indicate that reinforcement operated symmetrically under the conditions employed.

## DISCUSSION

In the kind of problem solving behavior investigated here the evidence is substantial of a gross inability to handle negative information. Neither conditions of reinforcement nor type of concept employed had any appreciable effect on such information handling. Instructions, manipulated so as to increase the flexibility with which a subject structured the task, similarly had no discernible impact on frequency of exclusion solutions. In addition to the implications of these results for concept learning generally, and for computer simulation of human problem solving processes in particular, an attempt will be made to suggest why exclusion solutions are so infrequent and under what circumstances they are likely to be used.

First, as mentioned earlier there are two somewhat different, but contingent, processes involved in the type of concept learning investigated here. One is the actual discrimination of the relevant cue(s) necessary for concept identification, while a second is the strategy employed in searching out and testing cue information. The essence of a strategy is that it is a knowledge process that when once learned can be used in a variety of problem situations. The basic cue discrimination is limited only to the specific problem. In the present investigation, with the exception of the transformation (choice of opposite instance), the same basic cue discrimination was required for both exclusion and control subjects, while the strategies of information handling were different. Consequently it was not the discrimination itself that made for difficulty in learning, but rather that the exclusion condition required a problem solving approach that has very low strength in most response repertoires. In addition, studies of reversal learning (Buss, 1956; Kendler & D'Amato, 1955) have demonstrated that subjects can perform in terms of an exclusion solution once they have learned the now-to-be excluded cue as a positive instance. This again suggests that the difficulty must lie in the cue search strategy employed in the acquisition phase.

In the present task the subject proceeded by discriminating some cue and testing its appropriateness by pressing a lever. If he was not positively reinforced he could simply discard that cue and resample, as Bruner et al. (1956) found many of their subjects did, or store the cue as an instance of what the concept was not. The difference would be between an almost blind trial-and-error strategy and a memory strategy that, by remembering what is incorrect, could progressively narrow the relevant space of cues. However, where the stimulus array is complex with a fairly large space of possible cues then an attempt at memory storage of negative instances is likely to be inefficient, if not impossible. (Analogous to the haystack searcher trying to remember each piece of straw under which the needle is *not*.) An important variable then, controlling the type of strategy that will be employed should be the subject's perception of the task as embodying a cue space within his memory capacity. When faced with a complex stimulus array the subjects adopt a strategy of proceeding by an almost blind trial-and-error strategy of positive identification (cues as to what the concept is) with only minimal activation of memory processes. However, a memory strategy requiring a subject to remember instances of what the concept is not is more likely to be used where a subject perceives the potential cue space to be small or where the properties of this space are specified for him by the experimenter as Hovland and Weiss (1953) have done. Campbell (1960) has recently proposed within a "framework of the comparative psychology of knowledge processes, a blind-variation-and-selective-retention process as fundamental to all inductive achievements." The present analysis makes a similar assumption, such that the two strategies above are problem solving heuristics that are likely "inductive achievements" in the human organisms' learning history. The use of exclusion solutions then is a distinctly secondary or "higher level" problem solving strategy, and because of the load it puts on memory processes it is likely to have a low probability of evocation. Only under the restricted conditions suggested above are exclusion solutions likely to occur. However, once a tentative positive identification of the relevant cue space has occurred then negative instances can be used to increase the precision of the defining attributes of the concept. In complex tasks negative instances serve principally to signal, when an error occurs, that the subject's current cue discrimination is incorrect or incomplete and are not likely to be used as direct sources of information until some reasonably well-defined positive representation of the solution has been acquired.

## REFERENCES

BRUNER, J. S., GOODNOW, JACQUELINE J., & AUSTIN, G. A. *A study of thinking*. New York: Wiley, 1956.

BUSS, A. H. Reversal and nonreversal shifts in concept formation with partial reinforcement eliminated. *J. exp. Psychol.*, 1956, **52**, 162–166.

CAHILL, H. E., & HOVLAND, C. I. The role of memory in the acquisition of concepts. *J. exp. Psychol.*, 1960, **59**, 137–144.

CAMPBELL, D. T. Blind variation and selective retention in creative thought as in other knowledge processes. *Psychol. Rev.*, 1960, **67**, 380–400.

DONALDSON, M. Positive and negative information in matching problems. *Brit. J. Psychol.*, 1959, **50**, 253–262.

FRIEBERGS, V., & TULVING, E. The effect of practice on utilization of information from positive and negative instances in concept identification. *Canad. J. Psychol.*, 1961, **15**, 101–106.

HOVLAND, C. I., & WEISS, W. Transmission of information concerning concepts through positive and negative instances. *J. exp. Psychol.*, 1953, **45**, 175–182.

KENDLER, H. H., & D'AMATO, M. F. A comparison of reversal shifts and nonreversal shifts in human concept formation behavior. *J. exp. Psychol.*, 1955, **49**, 165–174.

KENDLER, T. S. Concept formation. *Annu. Rev. Psychol.*, 1961, **12**, 447–472.

SMOKE, K. L. An objective study of concept formation. *Psychol. Monogr.*, 1932, **42**(4, Whole No. 191).

(Received December 26, 1962)

*Journal of Educational Psychology*
1963, Vol. 54, No. 3, 160–164

# EFFICIENCY OF INITIAL LEARNING AND TRANSFER BY INDIVIDUALS, PAIRS, AND QUADS[1]

HERBERT J. KLAUSMEIER, WILLIAM WIERSMA, AND CHESTER W. HARRIS

*University of Wisconsin*

The quads attained the concepts initially in significantly less time than cepts initially. Subsequently, 16 from the quads, 16 from the pairs, and 16 individuals attained a transfer concept after a delay of about 30 sec.; an equal number attained a transfer concept after a delay of 12 min. The quads attained the concepts initially in significantly less time than the pairs and the individuals; however, those who learned initially as individuals performed significantly better on the transfer problem than those who learned in pairs and quads.

How efficiently persons attain concepts and other outcomes in the cognitive domain when working as individuals, in comparison with working in groups, is of significance in schools as in other behavioral settings (Klausmeier, 1961). According to Lorge, Fox, Davitz, and Brenner (1958), the considerable experimentation on the quality of group and of individual performance had not yet produced conclusive results. Later, Fox and Lorge (1962) reported that the quality of decisions of groups was equal to that of individuals when the time for group activity was increased from 50 to 100 minutes; however, when only 50 minutes of time were available, the individuals performed significantly better than the groups. Duncan (1959) concluded in his survey of problem solving performances that the more efficient individuals usually perform better than the mean performance of groups. In a simple type of problem situation involving anagrams, Anderson (1961) reported that groups of three achieved higher than groups of two, and groups of two higher than individuals. Anderson also reported that the large majority of experiments showed group performances to be superior to performances of individuals.

Though the results are yet inconclusive concerning performances of individuals and groups of varying size in initial learning situations, it is surprising that social psychologists and others have not yet attempted to determine the effects of initial learning as an individual or a member of a group upon subsequent performance when working alone. A possible contributing factor to this lack of experimentation is the relatively large number of subjects involved, the difficulty of scheduling, and other practical matters.

Though the problems of experimentation with groups are considerable, it is important to attempt to secure answers, for in education, government, and industry, committees and other forms of group activity are widespread. In many classrooms, also, small group activities are used widely in order to achieve outcomes in the affective domain, such as attitudes and values, which can be attained best in small groups. Small groups are also arranged in order to acquire concepts, solve problems in mathematics and science, and achieve other outcomes in the cognitive domain. At present, no consistent explanations are forthcoming concerning why persons working in groups, in comparison with working as individuals, are more or less efficient on the same type of tasks.

---

[1] The research reported herein was supported initially by the Research Committee of the Graduate School of the University of Wisconsin and later in accordance with a contract between the University of Wisconsin and the United States Office of Education, Department of Health, Education, and Welfare.

In the experiment now reported, the purposes were to ascertain whether university students acquired concepts equally efficiently when initially working as individuals, in pairs, and in groups of four, and to ascertain whether they performed equally well on an immediate and a delay transfer problem when working as individuals.

## PROCEDURE

### *Subjects*

One hundred twenty-eight subjects were drawn randomly from classes in educational psychology in two blocks of 64 during two semesters. They were told that participation in the experiment was a requirement of the course and that the main results would be reported back to them toward the end of the semester. All the subjects were in the age range of 21–25; the median and modal age was 21.

### *Experimental Material*

A 128-card, ordered figural board was used. It was designed so that each card contained each of seven attributes and one of two defining characteristics, the total display incorporating the attributes and defining characteristics in all possible combinations thus: Seven attributes—number of borders, continuity of borders, number of figures, size of figures, texture of figures, color of figures, shape of figures; with their respective defining characteristics—one or two, solid or broken, one or two, small or large, solid or spotted, red or green, circle or ellipse.

One hundred twenty-eight concepts embedded on this display board can be defined by the information on each card singly, e.g., one solid border with one small solid red circle, two broken borders with two large spotted green ellipses. One single concept can be defined by the information on *all 128* cards —one or two, broken or solid borders with one or two, small or large, solid or spotted, red or green, circles or ellipses. A large number of concepts can be defined by combining one or more of the seven attributes with one or two of the defining characteristics.

In this experiment, the concepts to be attained had two relevant attributes in combination with one or more defining characteristics, e.g., broken borders with circular figures, two bordered small figures, solid red figures, one large figure, etc. There are 84 of these concepts of two relevant attributes contained on the 128-card board, and for each there are 32 positive instances, or exemplars, and 96 negative instances. Stated differently, each concept of two relevant attributes serves to categorize all 128 cards into two groups, one group of 32 exemplars of the concept and another group of 96 nonexemplars; the latter are all exemplars of the concept which can be defined by the five nonrelevant attributes. The 128 cards have been placed on the board in an ordered array of 8 rows and 16 columns.

The four concepts to be attained in the initial learning situation by each individual, pair, and quad were as follows: broken borders, circular figures; two borders, small figures; solid red figures; and one large figure. The concept used in the transfer problem was one-border elliptical figures.

### *Experimental Procedure*

Table 1 shows the design of the experiment. Assignment to the three treatment groups was random. Thirty-two subjects participated as individuals in the initial learning situation; subsequently 16 of these engaged in the transfer problem immediately, and 16 others after a 12-minute delay during which they read an

TABLE 1

DISTRIBUTION OF SUBJECTS BY TREATMENT GROUPS UNDER INITIAL LEARNING AND TRANSFER CONDITIONS

| Treatment group | Initial learning | Immediate transfer | Delay transfer |
|---|---|---|---|
| Individual | 32 | 16 | 16 |
| Pair | 32 | 16 | 16 |
| Quad | 64 | 16 | 16 |

article. Thirty-two participated initially as members of pairs; subsequently, 16 of these participated immediately in the transfer problem, and 16 others after a 12-minute delay. Sixty-four subjects participated initially as members of 16 quads; subsequently 16 of these engaged in the transfer problem immediately, and 16 others after a 12-minute delay. The 32 subjects from the quads who did not participate in the transfer phase were eliminated randomly. Similarly, all subjects were assigned randomly to the immediate and delay transfer problems.

In the initial learning situation, four concepts in sequence were attained by the subjects as individuals in pairs, or in quads. A 4 × 4 Latin square design was used in assigning each of the 4 sequences of concepts to the individuals, pairs, and quads. Each subject, whether having learning initially as an individual, in a pair, or in a quad, worked the same transfer problem.

The subjects appeared at the Learning Laboratory and, after a brief introduction with the experimenter, received the following instructions:

> This experiment is concerned with how people attain concepts. You are going to have an opportunity to attain several concepts. You will be shown a display of cards with various borders, figures, and colors. [The subject was then shown a 32-card display board. The two defining characteristics of each attribute were pointed out, i.e., one or two figures, large or small figures, solid or textured figures, red or green figures, and circular or elliptical figures.] Notice that each card is different. There are, however, a number of ways that the cards may be grouped so that they possess some common feature or features. For example, all cards with circular figures could be grouped together. Can you show me what cards on this display board would belong to such a group? [The experimenter waited until the subject had pointed out all the cards that had circular figures.] That is very good. We might also group together all the cards with red elliptical figures on them. Can you show me which cards would belong to this group? [Again the experimenter made sure the subject found all positive instances.] O.K., that is fine.
>
> When I said all red elliptical figures could be grouped together, I was stating a rule or a concept for classifying the display. In this experiment, your job will be to find out what rule I have used for grouping a series of similar cards.
>
> The procedure is as follows. At the beginning I will show you one card which is a member of the concept to be attained. Then you are to select other cards which you want to test as belonging to the same concept. You may select a card for testing by reading off the number you see right below it. Each time you test a card, I will tell you if it is a member of the concept. As you find out what cards do and do not belong to the concept, you can determine the concept. Whenever you think you know it, you may stop, and check it here. We use this form for this purpose. [The experimenter then showed the subject how to fill out a checkoff slip.] If the concept is correct, the problem is over. If it is not I will simply say "not correct," and you will continue testing cards until you again feel you know what the concept is. You may offer as many concepts as you like. The job, however, is to get the concept as quickly as possible.
>
> Do you have any questions?

The instructions were read to the individuals, pairs, and quads. To each pair and quad the additional instructions were given that they should talk freely with one another while attaining the concepts and that any subject could offer a card choice or a concept. No other instructions were given and no leader or spokesman for any pair or quad was designated.

Upon receiving the instructions, the subjects worked at the task according to the three treatment groups. The subjects who did not engage in the immediate transfer problem left the laboratory and were instructed to read an article. After 12 minutes they were brought back to the laboratory for the delay transfer problem. The four initial concepts and the immediate and delay concepts were attained in a period of 30–60 minutes, depending upon the efficiency of the subject.

## RESULTS

There were 32 subjects who learned initially as individuals, but only 16 pairs and 16 quads. Therefore, in the analysis of the initial learning, 16 individuals were eliminated randomly in order to have the same *N* in the analysis of variance for initial learning, immediate transfer, and delay transfer. For the same reason, 32 of the 64 subjects who

TABLE 2

MEAN TIME TO CRITERION IN MINUTES FOR INITIAL LEARNING AND TRANSFER ACCORDING TO TREATMENT GROUP

| Treatment group | Initial learning | Immediate transfer | Delay transfer |
|---|---|---|---|
| Individuals | 4.62 | 2.34 | 3.40 |
| Pairs | 3.89 | 4.82 | 3.57 |
| Quads | 3.27 | 4.62 | 4.01 |

learned initially as members of quads were eliminated randomly from participation in the transfer part of the experiment.

The mean time to attain the four concepts initially was computed for each individual, pair, and quad, yielding a total of 48 scores, 16 for individuals, 16 for pairs, and 16 for quads. In both the immediate and in the delay transfer condition, there were 16 who had learned initially as individuals, 16 in pairs, and 16 in quads.

In Table 2 are given the mean scores. As may be observed, the rank order in efficiency of performance in initial learning was quads (3.27 minutes), pairs (3.89), and individuals (4.62). However, in immediate transfer, the subjects who learned initially as individuals were most efficient (2.34), with time to criterion about half that of the subjects who now worked as individuals but had learned initially in pairs (4.82) and quads (4.62). In delay transfer the rank order of efficiency was individuals, pairs, and quads, but the differences were not so large as for immediate transfer.

Three one-way analyses of variance were computed on initial learning scores, immediate transfer, and delay transfer.

The *F* ratio shown in Table 3 for initial learning indicates that the difference among the three treatment groups was significant beyond the .01 level. Subsequent *T* tests of the initial learning scores between treatment groups showed the pairs to be significantly better than the individuals at the .10 level, the quads better than the pairs at the .10 level, and the quads more efficient than the individuals at the .01 level.

For immediate transfer, the difference among treatment groups was significant beyond the .05 level. *T* tests between sets of means showed that the subjects who had learned as individuals were significantly more efficient than those who had learned in pairs or in quads at the .05 and .01 levels, respectively. Although in the same relative rank order, the difference among treatment groups was not significant for the delay transfer condition.

### *Discussion*

An important methodological contribution of this experiment is the design

TABLE 3

RESULTS OF THREE ONE-WAY ANALYSES OF VARIANCE ON TIME TO CRITERION FOR INITIAL LEARNING AND TRANSFER CONDITIONS BY GROUP SIZE

| Source | Initial learning | | | Immediate transfer | | | Delay transfer | | |
|---|---|---|---|---|---|---|---|---|---|
| | *df* | *MS* | *F* | *df* | *MS* | *F* | *df* | *MS* | *F* |
| Between group | 2 | 7.71 | 6.42** | 2 | 30.40 | 4.72* | 2 | 1.60 | $< 1$ |
| Error | 45 | 1.20 | | 45 | 6.44 | | 45 | 4.95 | |

* $p \leq .05$.
** $p < .01$.

which enabled the testing of the effects of initial learning conditions upon subsequent performance. In future experiments designed to assess the quality of performances of individuals and of groups of varying size, the experimenters should be concerned not only with performance in the initial situation but also in a subsequent one. It would have been possible in the present experiment to have grouped the individuals into pairs and quads for the transfer problem instead of having those who learned initially in pairs and quads to work subsequently as individuals. The present design thus permits the determination of the effects of initial learning in groups of varying size upon subsequent performance as individuals, and vice versa.

One substantive result of the experiment is that pairs and quads perform more efficiently initially than do individuals; but subsequently, when working as individuals, those who learned originally as individuals perform significantly better than those who learned in pairs and quads. The first conclusion is in agreement with other investigations reported earlier in this article. The latter conclusion cannot be checked with other experiments, for no other reports were found in which the attempt was made to investigate this phenomenon. Although this conclusion may be applicable only to the specific experimental arrangements as described, it emphasizes clearly the conclusion of Lorge et al., (1958) and Duncan (1959) that the superiority of group performance in comparison with individual performance has not been fully established.

The crossover in efficiency of individuals and quads may be explained tentatively. In the initial learning situation, the pairs and quads secured a large amount of information in a relatively short period of time, analyzed the information correctly, recalled the information, and deduced the correct concepts. Pairs and quads accomplished this better collectively than did individuals working alone; however, not all the members of the pairs and quads learned well. When on their own and working individually on the transfer problems, they could not apply, for they had not learned. Working alone initially, the individuals often guessed incorrectly what the concept was without having complete information, or did not analyze the information correctly, or did not recall it well. However, each individual was active and most of them improved performance across concepts during the initial learning and also learned how to go about the task. They continued improving in the transfer situation by applying the learning-to-learn procedures in the transfer situation.

## REFERENCES

ANDERSON, N. H. Group performance in an anagram task. *J. soc. Psychol.*, 1961, 55, 67–75.

DUNCAN, C. P. Recent research on human problem solving. *Psychol. Bull.*, 1959, 56, 397–429.

FOX, D. J., & LORGE, I. The relative quality of decisions written by individuals and groups as the available time for problem solving is increased. *J. soc. Psychol.*, 1962, 57, 227–242.

KLAUSMEIER, H. J. *Learning and human abilities: Educational psychology.* New York: Harper & Row, 1961. Ch. 5.

LORGE, I., FOX, D., DAVITZ, J., & Brenner, M. A survey of studies contrasting the quality of group performance and individual performance: 1920–1957. *Psychol. Bull.*, 1958, 55, 337–372.

(Received January 7, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 3, 165–171

# EFFECTS OF ACCELERATING BRIGHT OLDER ELEMENTARY PUPILS:

## A FOLLOW UP[1]

HERBERT J. KLAUSMEIER
*University of Wisconsin*

A follow-up study was completed toward the end of the 5th grade to determine the effects of accelerating bright older pupils from 2nd to 4th grade after a 5-week summer session. The Ss were 1 group of 10 boys and 10 girls, above the median CA of 2nd graders prior to being accelerated; 2 groups of 20 nonaccelerated 5th graders of superior abilities, 1 above and 1 below the median CA of nonaccelerated 5th graders; and 2 groups of 20 nonaccelerated 5th graders of average abilities, 1 above and 1 below the median CA of nonaccelerated 5th graders. No unfavorable academic, social, emotional, or physical correlates of acceleration were found. A comparison of these results with other research on acceleration without skipping of content suggests that pupils of superior abilities should complete the baccalaureate degree no older than age 21.

Excellent nurturance of the emerging abilities and affective characteristics of all pupils is the primary task of the school. It is a difficult task because of the wide variability among the pupils in abilities and other characteristics and also because of relatively inflexible administrative arrangements which do not permit adaptation to individual differences. Some of these arrangements do not facilitate bringing together appropriate content and the abilities of pupils, yet efficiency of learning might be vastly improved through a relatively simple change in connection with acceleration.

Many school systems require children to have attained a chronological age of at least 5-9 as of September 1 in order to enter the first grade. They also require each child to spend at least 12 years completing 12 grades. In these school systems, the median age of entering first graders is 6-3, and the range is from 5-9 to 6-9. The pupils who spend 12 years completing the 12 grades are about 17-6 to 18-6 when graduating from high school in June and about 21-6 to 22-6 when graduating from college. The age at which most students are graduated depends upon the date of birth, not on any intellectual, educational, social, emotional, or physical characteristic.

There are as many bright and emotionally, socially, physically mature children among the older half in chronological age entering the first grade as there are below average, and emotionally, socially, and physically immature children among the lower half in chronological age. To keep an older bright child throughout his school life with younger, less mature children keeps him out of a productive career for at least 1 year, and it may also reduce his learning efficiency and motivation for learning.

Adherence to a schedule of one grade per year is no longer feasible since ac-

---

[1] This study was supported by the Wisconsin Improvement Program—Local School Systems and Teacher Education, and by the Research Committee of the Graduate School of the University of Wisconsin with funds from the Brittingham Trust.

celeration can be accomplished without skipping any essential content. However, the only controlled experiment reported thus far, dealing with the acceleration of bright, older, elementary school pupils, was by Klausmeier and Ripple (1962). The effects of accelerating bright, older pupils from the second to the fourth grade after a 5-week summer session were favorable toward the end of the fourth grade. In this article the effects of acceleration toward the end of the fifth grade are reported and discussed.

## METHOD

### *Subjects*

The subjects were 50 girls and 50 boys, 10 of each sex having been accelerated from the second to the fourth grade, and currently enrolled in the fifth grade. The other four groups of 10 girls and 10 boys each were two groups of fifth graders of superior abilities, one above and the other below the median age of fifth graders; and two groups of fifth graders of average abilities, one above and the other below the median age of fifth graders. Hereafter, the five groups are designated:

1. Acc—referring to the accelerates
2. 5SY—referring to the nonaccelerated pupils of superior ability below the median age of normally progressing fifth graders
3. 5SO—referring to the nonaccelerated pupils of superior ability above median age
4. 5AY—referring to the nonaccelerated pupils of average ability below median age
5. 5AO—referring to the nonaccelerated pupils of average ability above median age

Two other groups, originally third graders in the first year of experimentation, were not included in the follow up, one because about half of its members were accelerated from third to fifth grade and the second because appropriate standardized educational achievement testing could not be carried out during the follow up in the fourth grade.

As beginning fifth graders the mean chronological age of Acc was 9-6; the mean of 5SY and 5AY was 10-0, and the mean of 5AO and 5SO was 10-5. If no further acceleration is done the Acc will be graduated from high school in June at a mean age of 17-4, 5SO at a mean age 18-3, with the range from about 17-9 to 18-9.

### *Data Gathered and Treatment of Data*

Nine types of data were gathered on each pupil toward the end of the fifth grade, using the same instruments as were used toward the end of the fourth grade, except that different forms of standardized tests were used. The reliability and validity of the ratings and other scores were reported in some detail previously (Klausmeier & Ripple, 1962) and are not presented again.

A brief description of each instrument follows.

*Educational Achievements.* The Metropolitan Achievement Tests, Intermediate Battery-Partial were administered to all pupils. Scores were obtained for Word Knowledge, Reading, Spelling, Language Total, Language Study Skills, Arithmetic Computation, Arithmetic Problem Solving and Concepts, and Total Battery. The raw scores obtained were changed to standard scores, using the data in the manual accompanying the tests.

*Problem Solving Ability.* A test of 25 multiple-choice items was devised by the author and research assistant, Richard Ripple, now at Cornell University, to assess a composite problem solving ability as represented in items dealing with analogies, logical reasoning, pertinent questions, sensitivity to problems, and following directions. Each item correctly answered was scored 1.

*Creative Thinking Abilities.* Seven tests of creative thinking yielding 10 scores were administered to all pupils: Object Uses, Word Uses, Plot Titles, Expressional Fluency, Plot Questions, Object Improvement, and Sentence Improvement. These tests, except Sentence Improvement, were adapted from Guilford (Guilford, Kettner, & Christensen, 1956) by Frank B. May, Research Assistant, now at Washington State University, for use with children. Sentence Improvement was devised by May. The reliability coefficients of these instruments were reported previously (Klausmeier & Ripple, 1962); validity and other information about each instrument were presented by May (1961) and Ripple (1961).

*Psychomotor Abilities.* Four tests of psychomotor abilities were devised and administered by Grace Piskula, Physical Education Supervisor of the Racine public schools: Basketball Throw to measure strength and coordination of body and arms, Jump and Reach to measure coordination of arms and legs, Wall Pass to measure eye-hand coordination and speed of reaction, and Shuttle Run to measure running speed and maneuverability.

*Handwriting Skills.* A speed score was obtained by having each pupil write a standard sentence as rapidly as he could for 3 minutes.

The score was a count of the number of words written. A legibility score was obtained by having each pupil write the standard sentence once at his normal speed of handwriting. This sample was rated by two judges on a 9-point scale. The two ratings were summed and a high rating is indicative of high legibility. The procedures of Virgil E. Herrick, University of Wisconsin, for administering and scoring these tests were followed.

*Peer Acceptance.* A sociometric instrument was used in which the pupils listed their five best friends in rank order. A score for each pupil was obtained by tabulating the choices received, weighting the choices (5 for a first choice, 4 for a second choice, etc.), and adding the weighted scores.

*Attitudes toward School and Learning.* A ranking scale and an inventory were devised to assess the pupils' attitudes toward school and learning. The ranking scale, titled Places Liked Best, required the pupil to rank the school classroom and eight other places from highest (1) to lowest (9) on the basis of preference for the places. A low numerical ranking for the classroom indicates high preference for the classroom in relation to the other eight places: farm, school library, city park, school gym, neighborhood, public library, school playground, home.

The Attitudes Inventory consisted of 24 scaled items, all of which required the child to indicate his liking for something connected with school.

*Ethical Values.* A multiple-choice inventory was devised by the writer and research assistant, William Franzen (1961), now at Toledo University, to assess ethical values. Responses were keyed as to whether they indicated rational conscientiousness, irrational conscientiousness, group conformity, or self-centeredness. The inventory yielded a score for each of these four traits. A high score on rational conscientiousness was interpreted as being more desirable than a high score on any other trait. Item 1 of the inventory follows; the letters in parentheses following each choice indicate the keying:

Boys and girls are usually told not to run in the halls. You would

A. run in the hall to report a child injured on the playground. (RC)
B. run in the hall only when other boys and girls do. (GC)
C. never run in the hall because it is wrong to do so. (IC)
D. run in the hall in order to get to class on time. (SC)

*Intellectual and Affective Characteristics.* A Teacher Rating Scale was devised by Klausmeier and Ripple, consisting of 12 descriptive items. The teacher rated the pupil on each item, using a 4-point scale. Weights of 1 to 4 were given to each item and a total score was obtained by adding the weighted scores.

A $2 \times 5$ analysis of variance was run to determine the significance of difference among the means of the five groups. Where the difference was significant at the .05 level or beyond, a Tukey gap test was run to ascertain which groups were significantly higher or lower than Acc. Differences between other groups, such as 5AY and 5SO, are not presented in the interest of brevity and conciseness.

## RESULTS

In Table 1 the mean scores and standard deviation units are presented for the total boys and girls in each group, an indication of whether the $F$ ratio among the means of the five groups was significant at or beyond the .05 or the .01 level, and an indication of whether boys or girls were significantly higher.

The difference among the means for the five groups was significant for 25 measures but not for the following 8: Word Uses—fluency, Plot Titles—fluency, Basketball Throw, Handwriting Speed, liking for the school classroom, attitudes toward school and learning. group conforming values, and self-centered values. In general, 5AY and 5AO groups were lower than the other three groups on the 25 measures where the difference was significant. The sex × group interaction was significant at the .05 level for one variable, namely, teacher rating of intellectual, social, and emotional characteristics. Subsequent Tukey gap tests on this variable for each sex resulted in a significant difference among the means at the .01 level as noted in Table 1.

The girls were significantly higher than the boys in five measures: Plot Questions, Object Improvement, Handwriting Legibility, Rational Conscientious values, and teacher rating. The

TABLE 1

MEANS, STANDARD DEVIATIONS, AND SIGNIFICANT *F* RATIOS FOR 33 MEASURES

| Measure | 5SO | | Acc | | 5SY | | 5AO | | 5AY | | Significance of *F* Ratio | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | Groups | Sex |
| Metropolitan Achievement Tests | | | | | | | | | | | | |
| Word Knowledge | 56.60 | 7.47 | 63.45 | 6.46 | 57.00 | 4.65 | 50.05 | 5.13 | 47.95 | 5.09 | .01 | *ns* |
| Reading | 57.30 | 7.70 | 64.10 | 9.73 | 57.10 | 5.61 | 50.25 | 5.86 | 47.55 | 8.03 | .01 | *ns* |
| Spelling | 58.55 | 7.49 | 63.60 | 8.25 | 59.45 | 6.19 | 50.45 | 6.74 | 48.10 | 7.45 | .01 | *ns* |
| Language Total | 55.15 | 6.86 | 63.55 | 4.68 | 56.40 | 4.59 | 43.45 | 12.39 | 47.15 | 7.75 | .01 | *ns* |
| Language Study Skills | 57.75 | 9.92 | 65.40 | 11.92 | 60.95 | 8.73 | 52.00 | 6.08 | 49.40 | 8.32 | .01 | *ns* |
| Arithmetic Computation | 53.70 | 9.84 | 55.20 | 9.40 | 53.10 | 5.89 | 46.70 | 3.50 | 42.20 | 6.65 | .01 | *ns* |
| Arithmetic Problem Solving and Concepts | 57.30 | 6.85 | 60.00 | 8.40 | 55.85 | 5.43 | 49.10 | 4.76 | 46.35 | 5.61 | .01 | *ns* |
| Social Studies Skills | 56.60 | 10.51 | 61.10 | 10.43 | 57.95 | 7.38 | 50.80 | 6.69 | 48.70 | 10.36 | .01 | *ns* |
| Problem Solving | 19.05 | 2.58 | 20.30 | 2.34 | 17.75 | 2.95 | 14.40 | 3.66 | 13.80 | 3.93 | .01 | *ns* |
| Creative Thinking Battery | | | | | | | | | | | | |
| Expressional Fluency | 6.30 | 2.20 | 7.10 | 3.08 | 7.10 | 3.40 | 4.60 | 2.80 | 4.30 | 2.15 | .05 | *ns* |
| Plot Questions | 14.95 | 3.25 | 18.05 | 4.51 | 18.10 | 4.62 | 12.30 | 4.59 | 12.35 | 4.50 | .01 | .05G |
| Object Improvement | 5.65 | 2.81 | 6.65 | 3.03 | 5.60 | 2.04 | 3.50 | 1.54 | 4.30 | 2.32 | .01 | .05G |
| Sentence Improvement | 15.70 | 5.02 | 18.15 | 4.40 | 13.80 | 4.70 | 11.20 | 4.15 | 8.00 | 3.68 | .05 | *ns* |
| Object Uses—fluency | 17.65 | 5.87 | 19.75 | 6.21 | 19.25 | 6.70 | 14.75 | 3.92 | 14.25 | 4.76 | .05 | *ns* |
| Object Uses—flexibility | 6.50 | 2.96 | 6.90 | 4.05 | 5.60 | 2.70 | 4.50 | 3.02 | 4.45 | 2.54 | .05 | *ns* |
| Word Uses—fluency | 9.75 | 3.95 | 9.45 | 3.90 | 10.90 | 3.46 | 8.75 | 3.42 | 9.70 | 4.09 | *ns* | *ns* |
| Word Uses—flexibility | 6.45 | 1.36 | 6.65 | 1.79 | 6.20 | 1.28 | 5.40 | 1.54 | 5.10 | 1.48 | .05 | *ns* |
| Plot Titles—fluency | 5.15 | 2.37 | 5.60 | 2.52 | 6.35 | 2.16 | 4.60 | 1.70 | 5.25 | 3.16 | *ns* | *ns* |
| Plot Titles—cleverness | 2.35 | 1.76 | 4.25 | 2.84 | 3.40 | 2.37 | 1.80 | 1.61 | 2.70 | 2.05 | .05 | *ns* |
| Psychomotor Abilities | | | | | | | | | | | | |
| Shuttle Run | 7.05 | .89 | 7.60 | .68 | 7.25 | .64 | 7.15 | .93 | 6.60 | .99 | .01 | .01B |
| Wall Pass | 10.40 | 3.05 | 12.05 | 2.96 | 11.20 | 2.48 | 11.60 | 2.89 | 9.95 | 2.51 | .05 | .01B |
| Jump and Reach | 8.90 | 2.43 | 10.25 | 1.92 | 9.25 | 1.59 | 10.25 | 1.89 | 8.65 | 3.03 | .05 | .01B |
| Basketball Throw | 33.80 | 15.86 | 39.10 | 15.20 | 35.60 | 10.60 | 37.60 | 12.34 | 34.15 | 11.69 | *ns* | .01B |
| Handwriting Speed | 51.40 | 10.76 | 57.40 | 10.16 | 55.45 | 9.41 | 51.10 | 10.49 | 49.00 | 12.98 | *ns* | *ns* |
| Handwriting Legibility | 14.10 | 2.75 | 16.40 | 3.28 | 14.80 | 3.19 | 12.50 | 1.96 | 13.20 | 2.31 | .01 | .01G |
| Sociometric Choices | 12.65 | 9.17 | 21.40 | 11.65 | 14.25 | 9.41 | 16.40 | 12.12 | 9.05 | 6.95 | .01 | *ns* |
| Places Liked Best | 5.85 | 2.25 | 5.85 | 2.43 | 6.45 | 2.68 | 6.65 | 2.52 | 7.20 | 2.21 | *ns* | *ns* |
| Attitude Inventory | 52.85 | 6.38 | 55.70 | 5.14 | 53.85 | 6.53 | 50.55 | 8.31 | 50.40 | 8.04 | *ns* | *ns* |
| Ethical Values Inventory | | | | | | | | | | | | |
| Rational Conscientious | 10.80 | 2.46 | 11.35 | 2.72 | 9.85 | 3.10 | 7.75 | 2.75 | 8.05 | 3.36 | .01 | .05G |
| Irrational Conscientious | 2.95 | 1.99 | 2.80 | 1.67 | 3.70 | 2.43 | 5.10 | 2.36 | 4.50 | 3.14 | .01 | *ns* |
| Group Conforming | .45 | .83 | .10 | .31 | .45 | .69 | .55 | .60 | .40 | .68 | *ns* | .05B |
| Self-Centered | 1.80 | 1.61 | 1.70 | 1.38 | 2.00 | 1.38 | 2.55 | 2.11 | 3.00 | 1.75 | *ns* | *ns* |
| Teacher Rating | 34.50 | 4.42 | 38.80 | 4.58 | 32.50 | 5.66 | 26.80 | 5.98 | 24.85 | 5.38 | .01 | .01G |

boys were significantly higher than the girls in five measures: Shuttle Run, Wall Pass, Jump and Reach, Basketball Throw, and Group Conforming values. Possibly, no differences should have been expected between boys and girls in educational achievements since each child had to meet the same minimum criterion of educational achievement when originally selected for inclusion in the various groups.

In Table 2 are given the comparisons of Acc group with each other group on the 33 measures. In the first column of Table 2 are listed the groups which did not differ significantly from Acc on the various measures; in the second column are given the groups which were significantly lower than Acc; and in the third column are given the groups which were higher than Acc. Since the difference among the five groups was not significant on eight measures, Acc group does not differ from other groups on these eight measures. Further, Acc group was between 5SO and 5AY groups on some measures where the difference among the five groups was significant and thus may not be significantly different from any of the other four groups.

The results pertaining to the comparison of Acc group with each other group are now given, according to the main types of data gathered.

1. On eight scores of the Metropolitan Achievement Test, Acc group was not significantly different from the non-

TABLE 2

SUMMARY OF DIFFERENCES BETWEEN THE ACCELERATED AND THE FOUR COMPARISON GROUPS ON 33 MEASURES

| Measure | Not significantly different from Acc group | Significantly lower than Acc group | Significantly higher than Acc group |
|---|---|---|---|
| Metropolitan Achievement Test | | | |
| Word Knowledge | 5SY | 5AO, 5AY | 5SO |
| Reading | 5SO, 5SY | 5AO, 5AY | |
| Spelling | 5SO, 5SY | 5AO, 5AY | |
| Language Total | 5SY | 5AO, 5AY | 5SO |
| Language Study Skills | 5SO, 5SY, 5AO | 5AY | |
| Arithmetic Computation | 5SO, 5SY | 5AO, 5AY | |
| Problem Solving and Concepts | 5SO, 5SY | 5AO, 5AY | |
| Social Studies Skills | 5SO, 5SY, 5AO, 5AY | | |
| Problem Solving | 5SO, 5SY | 5AO, 5AY | |
| Creative Thinking Battery | | | |
| Expressional Fluency | 5SO, 5SY, 5AO, 5AY | | |
| Plot Questions | 5SO, 5SY, 5AO, 5AY | | |
| Object Improvement | 5SO, 5SY, 5AY | 5AO | |
| Sentence Improvement | 5SO, 5SY | 5AO, 5AY | |
| Object Uses—fluency | 5SO, 5SY, 5AO, 5AY | | |
| Object Uses—flexibility | 5SO, 5SY, 5AO, 5AY | | |
| Word Uses—fluency | 5SO, 5SY, 5AO, 5AY | | |
| Word Uses—flexibility | 5SO, 5SY, 5AO | 5AY | |
| Plot Titles—fluency | 5SO, 5SY, 5AO, 5AY | | |
| Plot Titles—cleverness | 5SO, 5SY, 5AO, 5AY | | |
| Psychomotor Abilities | | | |
| Shuttle Run | 5SO, 5SY, 5AO, 5AY | | |
| Wall Pass | 5SO, 5SY, 5AO, 5AY | | |
| Jump and Reach | 5SO, 5SY, 5AO, 5AY | | |
| Basketball Throw | 5SO, 5SY, 5AO, 5AY | | |
| Handwriting Speed | 5SO, 5SY, 5AO, 5AY | | |
| Handwriting Legibility | 5SY, 5AO, 5AY | | 5SO |
| Sociometric Choices | 5SO, 5SY, 5AO, 5AY | | |
| Places Liked Best | 5SO, 5SY, 5AO, 5AY | | |
| Attitude Inventory | 5SO, 5SY, 5AO, 5AY | | |
| Ethical Values Inventory | | | |
| Rational Conscientious | 5SO, 5SY | 5AO, 5AY | |
| Irrational Conscientious | 5SO, 5SY, 5AY | | 5AO |
| Group Conforming | 5SO, 5SY, 5AO, 5AY | | |
| Self-Centered | 5SO, 5SY, 5AO, 5AY | | |
| Teacher Rating | 5SO, 5SY | 5AO, 5AY | |

accelerated 5SY and 5SO groups, except in Word Knowledge and Language Total where 5SO was significantly higher; Acc group was significantly higher than the nonaccelerated 5AO and 5AY groups except in Language Study Skills, where 5AO group was not significantly different.

2. In Problem Solving as measured by a special test, Acc group was significantly higher than 5AO and 5AY groups and not significantly different from the nonaccelerated 5SO and 5SY groups.

3. On 10 tests of creative thinking abilities, Acc group was not significantly different from any other group, except on Object Improvement, Sentence Improvement, and Word Uses—flexibility; on these Acc group was higher than one or both 5AO and 5AY groups.

4. In six Psychomotor Abilities Acc group was not significantly different from any nonaccelerated group except in Handwriting Legibility, where 5SO group was significantly higher.

5. In number of choices received on a sociometric test, Acc group did not differ significantly from any other group.

6. In liking for school and learning, Acc group was not significantly different from any other group.

7. On four measures of ethical values, Acc group was not significantly different from the nonaccelerated 5SO and 5SY groups; it was significantly higher than 5AO and 5AY in Rational Conscientiousness and lower than 5AO in Irrational Conscientiousness.

8. In teacher ratings of the intellectual and affective characteristics of the pupils, where the sex × group interaction was significant, subsequent tests for each sex showed Acc boys and girls to be significantly higher than 5AO and 5AY boys and girls and not to differ significantly from 5SO and 5SY boys and girls.

In summary, toward the end of the fifth grade and the second year after acceleration, the accelerated pupils were equal to or surpassed the nonaccelerated pupils of average abilities and the younger pupils of superior abilities in all measures of intellectual and psychomotor abilities and adjustment; also, the accelerated pupils were significantly lower than nonaccelerated older bright pupils only in Word Knowledge, Language Total, and Handwriting Legibility.

## DISCUSSION

A principal objection to acceleration is that essential subject matter content is skipped, resulting in lower educational achievement of accelerated students. This objection was shown to be invalid in this and other experiments. In Table 3 are presented methods of acceleration which do not involve any skipping of content, the school level at which experimentation has been done, and indication of the reference in which the results are reported as favorable in terms of the educational achievements of the accelerates.

Other objections to acceleration revolve about the personal, social, and emotional adjustment of the accelerates. In Tables 1 and 2 of this article, substantial evidence is presented that accelerating bright older pupils, after attending school in the morning for 5 weeks during the summer, produces no undesirable effects on their personal, social, or emotional adjustment. The same conclusion was drawn in the other experiments cited in Table 3.

Since acceleration by one or two semesters without skipping may be accomplished at any school level with desirable results, it appears unwise to force bright students to spend 12 years in completing high school and 16 years in completing the requirements for the baccalaureate degree. Terman (1954) favored gifted students entering college not older than age 17. In terms of present shortages of teachers and other

TABLE 3

MEANS OF ACCELERATING PUPILS WITHOUT SKIPPING OF SUBJECT CONTENT

| Method | Level of schooling | Investigator |
|---|---|---|
| Being admitted early to kindergarten or first grade | Preschool | Hobson, 1948<br>Birch, 1954<br>Worcester, 1956 |
| Completing six semesters of the nongraded primary school in four or five semesters | Primary school | Klausmeier, 1962 |
| Attending summer session, mastering content of next higher grade, going into higher grade in fall | Elementary school | Klausmeier and Ripple, 1962 |
| Condensing or compressing content into shorter time periods, *e.g.*, 3 years of English into 2 years | Junior high school | Justman, 1954a<br>Justman, 1954b<br>Klausmeier, 1962 |
| Taking advanced college placement courses in the senior year of high school | Senior high school | Meister, 1956<br>Barnette, 1957 |
| Taking additional classes and summer session attendance | College | Flesher and Pressey, 1955 |

educated talent, an appropriate national educational goal might be to have all bright students complete the baccalaureate degree at least by their twenty-first birthday and the most able and rapidly maturing by age 20. Since we have compulsory public education of pupils of all abilities, excellent nurturance of the superior abilities of the top 5% or more is improved when appropriate subject content and emergent abilities are brought into harmony through acceleration.

## REFERENCES

BARNETTE, W. L. Advanced credit for the superior high school student. *J. higher Educ.*, 1957, **28**, 15–20.

BIRCH, J. W. Early school admission for mentally advanced children. *Except. Children*, 1954, **21**, 84–87.

FLESHER, MARIE A., & PRESSEY, S. L. Wartime accelerates ten years after. *J. educ. Psychol.*, 1955, **46**, 228–238.

FRANZEN, W. The effects of four prior and two current instructional arrangements upon the performance of students of superior learning ability. Unpublished doctoral dissertation, University of Wisconsin, 1961.

GUILFORD, J. P., KETTNER, N. W., & CHRISTENSEN, P. R. A factor analytic study across the domains of reasoning, creativity, and evaluation: II. Administration of tests and analysis of results. *U. Sth. Calif. Psychol. Lab. Rep.*, 1956, No. 16.

HOBSON, J. R. Mental age as a workable criterion for school admission. *Elem. Sch. J.*, 1948, **48**, 312–321.

JUSTMAN, J. Academic achievement of intellectually gifted accelerants and non-accelerants in junior high school. *Sch. Rev.*, 1954, **62**, 142–150. (a)

JUSTMAN, J. Academic achievement of intellectually gifted accelerants and non-accelerants in senior high school. *Sch. Rev.*, 1954, **62**, 469–473. (b)

KLAUSMEIER, H. J. Summary report of research completed, 1958–1961, on educational provisions for children and youth of superior learning abilities, Milwaukee Public Schools. Milwaukee: Board of School Directors, 1962.

KLAUSMEIER, H. J., & RIPPLE, R. Effects of accelerating bright older pupils from second to fourth grade. *J. educ. Psychol.*, 1962, **53**, 93–100.

MAY, F. B. Creative thinking: A factorial study of seventh-grade children. Unpublished doctoral dissertation, University of Wisconsin, 1961.

MEISTER, M. Cooperation of secondary schools and colleges in acceleration of gifted students. *J. educ. Sociol.*, 1956, **29**, 220–227.

RIPPLE, R. E. A controlled experiment in acceleration from the second to the fourth grade. Unpublished doctoral dissertation, University of Wisconsin, 1961.

TERMAN, L. M. The discovery and encouragement of exceptional talent. *Amer. Psychologist*, 1954, **9**, 221–230.

WORCESTER, D. A. *The education of children of above average mentality.* Lincoln: Univer. Nebraska Press, 1956.

(Received February 25, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 3, 172–176

# USE OF WRITTEN RESPONSES IN THE STIMULATED RECALL METHOD[1]

C. J. KRAUSKOPF[2]
*Ohio State University*

To determine whether written responses could be substituted for oral responses in the stimulated recall technique, 53 students from a course in general psychology were given a lecture on intelligence and personality defense mechanisms, and asked to write responses to a subsequent tape recorded presentation of the original lecture. The responses were rated for relevant thought and correlated with an immediate test and with final grades in the general psychology course. Correlations significant at the .01 or .05 level were obtained with both criteria. It was concluded that written responses were sufficiently reliable and valid to use in further investigation of classroom learning.

Most of the research in classroom learning has consisted of manipulation of the classroom situation and measurement of amount of material retained under various conditions, e.g., discussion versus lecture method of presentation. The method of stimulated recall appears to be a promising technique for investigating responses that occur in all classroom situations, but that are not easily measurable by the currently most frequently used techniques. The stimulated recall method elicits additional observable responses related to thought content and process and to degree of attention.

The technique of stimulated recall as used in the past by Bloom and others (Bloom, 1954; Gaier, 1950; Shulz, 1951), involves the following steps: (*a*) the tape recording of a classroom session, lecture, discussion, or similar material; (*b*) the playback of this tape in individual interviews, stopping the tape in several places and asking the students to state what they were thinking about at that point in the original class; (*c*) rating the responses as to the degree to which those thoughts were relevant to the lecture. Bloom (1953) classified the thoughts reported,

> in terms of their relationship to the topic or problem being discussed at the time, and an index was derived to represent the approximate proportion of the thoughts which were relevant [p. 167].

It appears that such ratings provide an indication of the nature and degree of covert class participation of students.

The research on learning in the classroom seems to be adequately summarized by Jacob (1957). Reference to previous use of the stimulated recall technique can be found in articles by Gaier (1952) and by Bloom (1953, 1954). The responses so obtained have been compared, in lecture and discussion, with observable behavior and with personality variables. The degree to which "thoughts" are found to be relevant to lecture material seems to be relatively independent of overt participation in class and to be a good predictor of course grades (Bloom, 1953).

The purpose of the present study is to determine the effectiveness of a group method of collecting thought responses in which subjects were asked to write down their responses to the original presentation. If effective, this procedure would permit more efficient use of research time in collecting data, and more extensive investigation of classroom learning might be carried out in a manner not possible with the more common techniques involving extensive individual interviewing.

---

[1] Based on the author's doctoral dissertation (Krauskopf, 1960).

[2] Now at the University of Missouri.

## METHOD

This experiment was conducted with 54 students, mainly second quarter freshmen, in an introductory psychology course. In ability they approximate the general distribution of freshmen on the Ohio State Psychological Examination (OSPE). The $\chi^2$ test between the subjects' OSPE scores and the distribution of the freshman class yields a probability of less than .30 that there is a difference between the distributions. One student failed to complete the course and was dropped from the study.

### *Hypotheses*

The major hypotheses to be tested, stated as null hypotheses, follow:

1. There is no difference between students high on relevant thought and students low on relevant thought during a lecture on general psychology in performance on a test of application of principles and interpretation of data covering material from general psychology.

2. There is no difference between students high on relevant thought and low on relevant thought during a lecture on general psychology in their final grade for the course.

It was predicted that the null hypotheses would be rejected. The .05 level of probability was selected as a criterion for rejection of the null hypotheses.

### *Procedure*

A lecture consisting of material from the general psychology course, but which had not yet been covered in the subjects' class, was memorized and presented by the experimenter to each group of students. This lecture was taped and then played back an hour later. The students were instructed to follow the tape as they had the original lecture, and when the tape stopped, to write down whatever they happened to be thinking at that moment, not to worry about complete sentences, but to get down quickly whatever they were thinking, whether it was on the topic of the lecture or not. The tape was stopped at seven critical points, where a student could draw a conclusion or synthesize material, and at eight other predetermined points.

An immediate test was given consisting of application of principles and interpretation of data items. The items were drawn from material covered earlier in the general psychology course, or were answerable from the data presented.

The responses obtained during the playback were then rated by two judges independently for relevant thought. The ratings were made by classifying the students' papers into six categories, from highly relevant, active use of outside material in the first category to degree of irrelevant material and incorrect information in the sixth category. Instructions to the judges were to rate the papers on relevance of material, references to self- and personal experience, assigning Number 1 to the best papers through Number 6 to the worst. Almost pure passive listening was to be rated about the middle. Wrong information and misinterpretation of material presented was to be rated low.

These ratings were intercorrelated, using the Pearson $r$, with the score on the immediate test, the final course grade in general psychology, and OSPE score. In order to estimate the overlap of the relevant thought ratings with a standard predictor of course grades, partial correlations were computed between the ratings and both the immediate test score and final grade with OSPE held constant. Multiple correlations were also computed in which the relevant thought rating and OSPE score were used to predict the test and course grade.

The ratings used in the computations were combined numerical ratings of the two judges. The correlation between the ratings of the two judges was .77.

## RESULTS

Intercorrelations among the relevant thought ratings, OSPE scores, immediate test scores, and final grades in general psychology were computed and are presented in Table 1.

The means and standard deviations of the various scores are presented in Table 2.

TABLE 1

INTERCORRELATIONS OF RELEVANT THOUGHT RATING OSPE, TEST SCORE, AND FINAL GRADE IN GENERAL PSYCHOLOGY

| | Relevant thought rating | OSPE | Test | Final grade |
|---|---|---|---|---|
| Relevant thought rating | | .08 | .56** | .27* |
| OSPE | .08 | | .30* | .65** |
| Test | .56** | .30* | | .34* |
| Final grade | .27* | .65** | .34* | |

* $p \leq .05$.
** $p \leq .01$.

These results indicate that the hypothesis of no difference between students high and low on relevant thought on an immediate test of application of principles and interpretation of data items can be rejected at the .01 level of probability and that the hypothesis of no difference in final grades for the general psychology course can be rejected at the .05 level.

The partial correlation between relevant thought rating and immediate test score with OSPE constant is .60. The partial correlation between relevant thought rating and final grade, with OSPE constant is .31. The multiple correlation of relevant thought rating and OSPE with immediate test score is .61. The multiple correlation of relevant thought rating and OSPE with final grade is .68.

TABLE 2

MEANS AND STANDARD DEVIATIONS OF RELEVANT THOUGHT RATING, OSPE, TEST SCORE AND FINAL GRADE IN GENERAL PSYCHOLOGY

| Statistic | Relevant thought rating (combined) | OSPE | Test | Final grade |
|---|---|---|---|---|
| *M* | 3.86 | 56.0 | 20.6 | 209 |
| *SD* | 1.06 | 24.1 | 2.93 | 35.6 |

## DISCUSSION

### *Academic Prediction with Written Responses*

Table 1 indicates that the two major null hypotheses should be rejected. Other criteria are needed to evaluate the usefulness of written responses as a research tool. In order to be maximally useful a new predictor should add to or improve on other standard predictors of academic performance, such as OSPE. The partial correlation of the relevant thought ratings and the immediate test with OSPE held constant indicates that the ratings add substantially to the prediction of application of principles and interpretation of data test items. The partial correlation of the relevant thought ratings and the final course grade with OSPE constant indicates that this might add something to the prediction of grades in general psychology. The multiple correlations of .61 and .68 for the combination of relevant thought and OSPE with immediate test and final grade add to predictive efficiency in both cases.

The correlation of OSPE and the relevant thought ratings is low enough to suggest that something is being measured other than what is usually measured by the standard predictors, e.g., OSPE.

Bloom (1953) reported greater difficulty in predicting grades from lecture material than from discussion, but he gives only one correlation coefficient. In this study there was no discussion by subjects during the original presentation, but in a pilot study where there was discussion by the subjects many of the questions by students seemed to focus the attention of other students. Additional research seems indicated using discussion classes and obtaining written responses.

### *Possible Explanation*

This method provides information otherwise difficult to obtain about attention. The attention attracting power of some phrases is very high. For example, "Where does little Pierre belong in school?" (in reference to the history of the Stanford-Binet) received the attention of nearly every student. Outside stimuli are occasionally very attractive. An airplane passing close overhead caught the attention of most. Unusual objects in the room were attractive. The room in which the experiment was conducted had electrical receptacles hanging from the ceiling, a fact noted by many of the students at some time during the lecture. Other examples could be given, but these seem sufficient to illustrate the point that attention is an important factor in making relevant thought ratings. An observer has no way of differentiating the student whose attention is directed toward the instructor from the student who is looking at the instructor and planning the next dance for his social group. Stimulated recall provides a method of making this differentiation.

According to Bloom (1954) the percentage of time during which a student's attention is directed toward subject matter is greater during discussion. In both Bloom's study and the present study the stimulated recall responses indicate qualitative differences in attention. These qualitative differences seem to be on an active-passive dimension. Inattention is indicated by responses as, "I'd like to be out with Mr.—[her psychology instructor] right now." At the passive end of the dimension are such responses as, "just listening" or repeating the last statement in the lecture. Others appear to spend much of the time applying information to people they know, e.g., "That sounds like my roommate." And still others make more general applications, "I wonder how many feelings are not hurt because of repression."

As the papers are evaluated the first factor considered is the ratio of statements indicating attention to those indicating inattention, a quantitative consideration of attention. The second factor considered is a qualitative difference in the statements indicating attention. Are these active or passive?

A third factor is the way in which a student considers outside material, information which was not present in the lecture presented. For example, "We studied something like this in sociology class. The instructor said intelligence was hard to determine." Or, "Fred is an example of this. He daydreams all the time." This type of response is not present except in those papers which would be evaluated as having active attention. After the amount and quality of attention are used to differentiate the responses, a further differentiation is made on the basis of integration of outside material.

Bruner, Goodnow, and Austin (1956) define *categorizing* as, "Additional inferences about other unobserved properties of the object or event [p. 244]." Allport (1954) mentions categorizing as classifying perceptual evidence. A person then infers other characteristics of a stimulus from the class to which it is assigned. Fragments of this categorizing process are apparent in the stimulated recall responses. For example, in reference to M. Itard in the story of the *Wild Boy of Aveyron* (Valentine & Wickens, 1949), "Must be a psychologist, sounds like one." To this student, one who takes on this type of training task is a psychologist. In another case a student finds an example for a newly acquired category, "Conversion behavior—my roommate making herself have a headache."

In retrospect the evaluation of the responses was made by judging the amount and quality of attention, the integration of outside material, and the amount and quality of categorizing.

### *Suggested Research*

It has been possible to reliably rate stimulated recall material for relevant thought with both individual interviews (Bloom, 1953) and, in this experiment, written responses; however, it would be useful to have data on the reliability of the technique in a test-retest situation. Also, if a better scoring system could be devised, it would be interesting to know whether there were any differences in the relevant thought stimulated by widely varying types of subject matter, e.g., physical sciences versus humanities. Bloom (1954) has identified differences resulting from use of lecture and discussion methods of presentation, but some individual students may show more relevant thought in a class where the scientific method of inquiry is emphasized and others in a class where the historical method is emphasized.

Courses in logic often claim as an aim of the course to teach students a different method of thinking. The Ohio State University College of Law is able to state definite ways in which they wish to change the thought processes of their students. If a scoring system can be devised based on the methods of categorizing information which are desired, it may be possible to investigate changes in thought processes through periodic administrations of this technique throughout the course. It may also be possible to identify the activities which produce a high proportion of tangential thought.

## REFERENCES

ALLPORT, G. W. *The nature of prejudice.* Cambridge: Addison-Wesley, 1954.

BLOOM, B. S. Thought processes in lectures and discussions. *J. gen. Educ.,* 1953, **7**, 160–169.

BLOOM, B. S. Thought processes of students in discussion. In S. J. French (Ed.), *Accent on teaching.* New York: Harper, 1954. Pp. 23–46.

BRUNER, J. S., GOODNOW, JACQUELINE J., & AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

GAIER, E. L. A study of memory under conditions of stimulated recall. *Amer. Psychologist,* 1950, **5**, 343. (Abstract)

GAIER, E. L. Selected personality variables and the learning process. *Psychol. Monogr.,* 1952, 66(17, Whole No. 349).

JACOB, P. E. *Changing values in college.* New York: Harper, 1957.

KRAUSKOPF, C. J. The use of written responses in the stimulated recall method. Unpublished doctoral dissertation, Ohio State University, 1960.

SHULZ, STELLA B. A study of relationships between overt verbal behavior and conscious mental processes of students. Unpublished doctoral dissertation, University of Chicago, 1951.

STERN, G. G., STEIN, M. I., & BLOOM, B. S. *Methods in personality assessment.* Glencoe, Ill.: Free Press, 1956.

VALENTINE, W. L., & WICKENS, D. D. *Experimental foundations of general psychology.* New York: Rinehart, 1949.

(Received March 1, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 4, 177–182

# CONCEPT ATTAINMENT AND ADOLESCENT DEVELOPMENT[1]

LEE YUDIN[2] AND SOLIS L. KATES
*University of Massachusetts*

The purpose of this experiment was to determine if concept attainment efficiency and strategies were related to adolescent development and method of presenting information. There were 40 Ss at the 12-yr., 14-yr., and 16-yr., levels, respectively, all of average intelligence. 4 groups of 30 Ss, 10 from each age level, solved 6 concept problems. Each group was given a different presentation of information. The 1st hypothesis was supported. The 14- and 16-year-old Ss were not different from each other and were superior to the 12-year-old Ss on efficiency and strategies. The 2nd hypothesis was not supported. Positive or negative series, transmitting equal information, appeared to have no different effects upon efficiency and strategies.

The purpose of this study was to determine if efficiency and strategy in concept attainment are related to age and to experimental method of presenting instances of the concept. The developmental theory of Inhelder and Piaget (1958) indicated that, as adolescents grow older, they develop greater facility in generalizing from the data at hand, in combining operations such as classification and seriation, in formulating and manipulating propositions, and in reasoning in a hypothetico-deductive fashion. Further experimental evidence is necessary to demonstrate that more efficient intellectual operations are used by 14-year-old adolescents as compared with 12-year-old adolescents, as suggested by these investigators.

Hovland and Weiss (1953) have explored whether negative instances or positive instances that transmit equal information are equally effective in concept attainment. Concept attainment, they found, was more efficient with positive instances than with negative instances. A pertinent question to be investigated is whether this finding applies to concept attainment by adolescents under optimal conditions of presenting information. Cahill and Hovland (1960) have demonstrated that a simultaneous successive method for concept attainment was superior to a successive method. One of their measures, the perceptual error—a guess about the concept that is incompatible with the information in the current instance—was found to be highly correlated with concept attainment efficiency.

In their concept attainment studies, Bruner, Goodnow, and Austin (1956) controlled the presentation of the instances and identified the instances as exemplars or nonexemplars of the concept. Their subjects guessed the concept after each instance and tended to follow ideal rules for maintaining or changing their guesses. Upon the basis of the Bruner et al. research, we have developed two methods of examining the sequence of guesses, the consistent and the ideal strategies. The consistent strategy indicates that the guess is maintained

---

[1] This study was supported, in part, by Research Grant M-4392 from the National Institutes of Health, United States Public Health Service and, in part, by United States Public Health Service Predoctoral Fellowship Grant No. 17,429 from the National Institutes of Health.

[2] Based upon a master's thesis completed by the senior author under the direction of the junior author.

or changed when the current instance is confirming or infirming. The ideal strategy is one where the guess is maintained or changed in accordance with the information presented by the current and all previous instances.

Perceptual error, consistent strategy, and ideal strategy, in that order, appear to be measures representing increasing maturity of cognitive operations. Perceptual error suggests an inability to deal with information given directly and perceptually. Consistent strategy shows an ability to deal with information by maintaining or changing guesses but without the systematic integration of previous information. Ideal strategy includes the ability defined by the consistent strategy and the additional ability to integrate systematically previous information.

The hypotheses are as follows:

1. Twelve-year-old subjects will require significantly more instances for concept attainment, will have significantly fewer consistent and ideal strategies, and will have significantly more perceptual errors than 14- and 16-year-old subjects. The 14- and the 16-year-old subjects will not differ significantly on these variables.

2. Concept attainment on positive instances will require significantly fewer instances, will have significantly more consistent and ideal strategies, and will have significantly fewer perceptual errors than concept attainment on negative instances presenting equal information.

## METHOD

### *Subjects*

There were 40 boys at the 12-year level, 40 boys at the 14-year level, and 40 boys at the 16-year level, making a total of 120 subjects. These boys were within 6 months of their birthdays. Subjects were from different public schools in western Massachusetts but were equally represented in the experimental groups to be described below.

### *Test Materials*

All the subjects were administered two practice problems and six experimental problems. Each problem consisted of 13 $2 \times 2$ inch slides. After the first slide, each successive cycle of 3 slides provided sufficient information to solve the problem. The problem was solved when the correct answer was offered twice consecutively without any subsequent incorrect responses.

The first slide in each problem presented one positive instance of the concept, the focus instance. The remaining 12 slides for any problem included, on each slide, the focus instance and one other instance. For example, on the slides of the positive series, the focus instance and one other positive instance of the concept appeared. On the slides of the negative series, the focus instance and a negative instance were shown.

Each instance showed a combination of four attributes: (*a*) one of three forms (circles, crosses, or squares), (*b*) one of three colors for the forms (red, black, or green), (*c*) one of three numbers of forms (one, two, or three), and (*d*) one of three borders (one, two, or three lines surrounding the forms). Several examples of instances are: (*a*) one red circle, two borders; (*b*) two black crosses, three borders; (*c*) three green squares, one border; (*d*) two black circles, one border.

The concepts to be solved were the same for the corresponding problems in the positive and negative series. For example, in both positive and negative series, the concept to be attained in the first problem was the color black; the focus instance in the first problem of the two series was three black squares and two borders. The second problem for both positive and negative series required all the subjects to discover that the concept was one form given the focus instance of one red cross and two borders. The same conditions obtained for the other four experimental problems (Yudin, 1962).

The instructions indicated that the correct concept in each problem was a single attribute. As mentioned above, the single attribute black was the concept for the first problem and the single attribute of one form was the concept for the second problem. Black crosses or one square could not be correct because these responses contained two attributes.

Equal information was presented by the corresponding slides of the corresponding problems in the positive and negative series. The second slide of the particular problem, whether in the positive or negative series, eliminated the same attribute. The third slide in this problem eliminated an additional but same attribute. For example, the second slide of the first problem, whether in the positive or negative series, eliminated number; the third slide of this problem eliminated form. The experimental problems, the instances contained on the 13 slides for each problem, and the attributes eliminated by each slide, are available in Yudin (1962).

### *Procedure*

The 120 subjects were divided into four experimental groups of 30 subjects each. The 30 subjects in any one experimental group included 10 12-year-olds, 10 14-year-olds, and 10 16-year-olds. The 12-year-old subjects in all the four experimental groups were matched on intelligence and educational achievement; similarly, the 14- and 16-year-old subjects in all four groups were matched on intelligence and educational achievement.

The first experimental group was shown positive instances in their six experimental problems. The second group was shown positive instances in the first three problems and negative instances in the second three problems. The third group solved the first three problems given negative instances and the second three problems given positive instances. The six problems solved by the fourth group contained negative instances.

The problems were presented to groups of five subjects. The attributes of the instances were pointed out and examples of instances were demonstrated. The full instructions are given in Yudin (1962). The experimental problems were administered when the subjects in each group had successfully solved the practice problems. All subjects solved the same six experimental problems.

We emphasized that any one attribute could be the correct concept for any problem. By this instruction, we tried to break any possible set that may have been established by the practice problems or by the preceding experimental problems. After each slide was exposed for 10 seconds, the subjects wrote their guesses about the concept for that particular problem. Each slide was exposed only once. The subjects covered their guesses with a sheet of paper provided by the experimenter.

For the number of instances, we noted the slide at which the correct concept was offered in each problem. The consistent strategies were determined by counting the number of times a previous guess was maintained when the new instance confirmed it or was changed to accord with the new instance when it was infirmed. For the number of ideal strategies, we counted the number of guesses that was maintained or changed when indicated by the new instance and by all previous instances. A perceptual error was counted when the guess was not in accord with the information given by the immediately preceding slide.

## RESULTS

The three age groups did not differ significantly on intelligence, all groups having an average IQ of 101 on the Otis Beta test. The four experimental groups did not differ significantly on age, intelligence, and school achievement.

### *Age and Concept Attainment (First Hypothesis)*

There was a significant difference in the number of instances required for concept attainment when the three age groups were compared (see Table 1). Further analysis by the Duncan range test (Edwards, 1960) showed that the 12-year-old subjects required significantly more instances than either the 14- or 16-year-old subjects. The differences between the mean number of instances for the 12-year-old group (54.8) and the 14-year- (44.1) and the 16-year- (42.9) old groups were 10.7 and 11.9, respectively (see Table 2). These differences exceeded the values of 7.2 and 7.7 required for significance at the .05 level. The difference (1.2) between the 14- and 16-year-old subjects was not significant for instances.

The results of the analysis of variance indicated that the main effect for the three age groups was significant on consistent strategies (see Table 1). Duncan range comparisons showed that the mean number of consistent strategies for the 12-year-old group (56.3) was significantly less than the mean number for the 14-year-old group (61.7) and the mean number for the 16-year-old group (63.4) where differences greater than

TABLE 1

ANALYSIS OF VARIANCE TABLE FOR DEPENDENT VARIABLES

| Source | *df* | Number of instances *SS* | Number of instances *F* | Consistent strategies *SS* | Consistent strategies *F* | Ideal strategies *SS* | Ideal strategies *F* | Perceptual errors *SS* | Perceptual errors *F* |
|---|---|---|---|---|---|---|---|---|---|
| Age (A) | 2 | 3425 | 5.9** | 1099 | 6.2** | 4838 | 8.5** | 500 | 4.2* |
| Presentation (B) | 3 | 527 | .6 | 189 | .7 | 1067 | 1.2 | 701 | 3.9* |
| A × B | 6 | 1763 | 1.0 | 327 | .6 | 2089 | 1.2 | 730 | 2.0 |
| Subjects/A × B | 108 | 31499 | | 9521 | | 30601 | | 6465 | |
| Total | 119 | 37215 | | 11136 | | 38596 | | 8396 | |

* $p \leq .05$.
** $p \leq .01$.

4.2 and 4.4, respectively, are significant at the .05 level (see Table 2). There was no significant difference between 14- and 16-year-old subjects on consistent strategies.

On the variable of ideal strategies, there was a significant finding when the three age groups were compared (see Table 1). Duncan range comparisons revealed that the differences in ideal strategies of 11.7 and 14.7 between 12-year-old (31.0) and 14-year-old (42.7) subjects and 12-year-old and 16-year-old (45.7) subjects, respectively, exceeded the differences necessary for significance at the .05 level (7.5 and 8.0) (see Table 2). No significant difference was found between the 14- and the 16-year-old subjects on ideal strategies.

A significant finding for perceptual errors indicated that the three age groups were different (see Table 1). The respective differences of 3.7 and 4.7 between the means for 12-year-olds (11.5) and 14-year-olds (7.8) and 12-year-olds and 16-year-olds (6.8) were significant where differences of 3.3 and 3.5 are required for significance at the .05 level. There was no significant difference between 14- and 16-year-old subjects on perceptual error.

These results supported all parts of the first hypothesis stating that 12-year-old subjects would be significantly inferior on all measures to 14- and 16-year-old subjects who would not be different.

### *Method of Presentation (Second Hypothesis)*

There were no significant differences on the number of instances, and on consistent and ideal strategies when per-

TABLE 2

MEANS AND STANDARD DEVIATIONS FOR AGE GROUPS ON DEPENDENT VARIABLES

| Age group | Number of instances *M* | Number of instances *SD* | Consistent strategies *M* | Consistent strategies *SD* | Ideal strategies *M* | Ideal strategies *SD* | Perceptual errors *M* | Perceptual errors *SD* |
|---|---|---|---|---|---|---|---|---|
| 12 years | 54.8 | 16.4 | 56.3 | 9.8 | 31.0 | 14.5 | 11.5 | 7.3 |
| 14 years | 44.1 | 17.0 | 61.7 | 9.3 | 42.7 | 17.7 | 7.8 | 7.5 |
| 16 years | 42.9 | 15.2 | 63.4 | 7.3 | 45.7 | 13.8 | 6.8 | 7.3 |

TABLE 3

MEANS AND STANDARD DEVIATIONS FOR METHOD OF PRESENTATION ON DEPENDENT VARIABLES

| Method of presentation | Number of instances | | Consistent strategies | | Ideal strategies | | Perceptual errors | |
|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| 6 positive | 43.3 | 17.5 | 61.4 | 10.5 | 41.9 | 17.7 | 5.1 | 5.8 |
| 3 positive plus 3 negative | 48.8 | 17.7 | 61.2 | 9.2 | 37.9 | 17.8 | 8.3 | 9.2 |
| 3 negative plus 3 positive | 49.6 | 13.1 | 58.3 | 8.5 | 39.3 | 11.9 | 12.2 | 6.7 |
| 6 negative | 46.3 | 16.1 | 61.0 | 7.1 | 40.1 | 15.6 | 8.9 | 7.3 |

formance on the positive series was compared with the negative series (see Tables 1 and 3). On the variable of perceptual errors, there was a significant finding (see Table 1). However, Duncan range analyses indicated that the number of perceptual errors made on the positive series (5.1) was not significantly different from the number of perceptual errors on the negative series (8.9; see Table 3). A difference of 4.2 was necessary for significance at the .05 level. Parenthetically, the significant finding in the analysis of variance was due to the number of perceptual errors (12.2) made in the negative-positive series. The second hypothesis was not confirmed; there were no differences on instances, consistent and ideal strategies, and on perceptual errors when the positive series was compared with the negative series.

## DISCUSSION

### *Age and Concept Attainment*

The first hypothesis was supported by the results; the 14- and the 16-year-old subjects were significantly superior to the 12-year-old subjects in concept attainment efficiency. The strategies employed by the 14- and the 16-year-old subjects were also superior to those of the 12-year-old subjects. The older groups did not differ either in efficiency or in strategies. These results confirm the description of adolescent development advanced by Inhelder and Piaget (1958). They suggest that the stage of formal reasoning operations begins at about 12 years and reaches equilibrium at about 14–16 years of age. At these latter ages, Inhelder and Piaget (1958) indicate that subjects succeed more often than 12-year-old subjects in solving problems demanding formal operations.

In the conceptual problems of this research, two rules for transforming information were used. First, positive instances have the correct attribute and second, negative instances do not have the correct attribute. Subjects had to infer that any attribute that did not appear in the positive instances would not be the correct concept. Some subjects had to make the additional inference that any attribute present in both focus and negative instances would not be the correct concept. In addition to these mental operations, the subjects had to remember the information supplied by the former instances. Our results then demonstrate that the 14- and the 16-year-old subjects are more capable in these mental operations than the 12-year-old subjects.

The results of our strategy variables can help us to determine how the 12-year-old subjects were inferior to the 14- and the 16-year-old subjects. Since the 12-year-old subjects made significantly more perceptual errors, it is evident that they are less efficient than the 14- and the 16-year-old subjects in dealing with information given directly and perceptually. Twelve-year-old subjects were more liable to specify, as the correct concept, an attribute that did not appear in the currently shown positive instance or to specify an attribute that did appear in the currently shown negative instance. The proper application of the two rules for transforming information presented by positive and negative instances was more difficult for the 12-year-olds than for the 14- and the 16-year-olds.

The 14- and the 16-year-old subjects showed significantly more consistent and ideal strategies than the 12-year-old subjects. Since the ideal strategies are counted as part of the consistent strategies and thus partially overlap, we will not separate these strategies in our discussion. The following are the inferences based upon the results with consistent and ideal strategies: (*a*) the older subjects were more capable of remembering their previous guesses to check them with the current instances; (*b*) they more readily maintained their guesses when the current instances were confirming or more readily changed them when the current instances were not confirming; (*c*) in the ideal strategies, they more readily maintained or changed these guesses to accord with all the previous instances.

### *Method of Presentation*

The results did not support the second hypothesis and demonstrated that positive or negative series, both presenting equal information, made very little difference in the number of instances required or in the strategies pursued for solution. In this experiment, unlike the Hovland and Weiss (1953) study, the negative and the positive instances were always presented with the focus (positive) instance. For the subjects who worked with the negative series, the focus instance was shown with each negative instance on each slide. On each slide of the negative series, the subjects were thus aware of the attributes that could be correct (in the focus instance) and of the one attribute that was not correct (in the negative instance). This presentation of negative instances in conjunction with the same focus instance may have resulted in relatively greater efficiency of concept attainment. Possibly, Smoke's (1932) original suggestion that negative instances might possibly contribute to efficiency of concept attainment holds only when each negative instance is shown together with the same positive instance of the concept. Because there were no redundant instances in each cycle, a more systematic elimination of incorrect attributes was possible without the added strain of dealing with redundant information. Finally, single attributes were the correct concepts, possibly reducing the level of difficulty in assimilating negative information to that of positive information.

## REFERENCES

BRUNER, J. S., GOODNOW, JACQUELINE, J., & AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

CAHILL, H. E., & HOVLAND, C. I. The role of memory in the acquisition of concepts. *J. exp. Psychol.*, 1960, **59**, 137–144.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Holt, Rinehart & Winston, 1960.

HOVLAND, C. I., & WEISS, W. Transmission of information concerning concepts through positive and negative instances. *J. exp. Psychol.*, 1953, **45**, 175–182.

INHELDER, BÄRBEL, & PIAGET, J. *The growth of logical thinking.* New York: Basic Books, 1958.

SMOKE, K. L. An objective study of concept formation. *Psychol. Monogr.*, 1932, **42**(4, Whole No. 191).

YUDIN, L. W. A developmental study of concept attainment. Unpublished master's thesis, University of Massachusetts, 1962.

(Received March 28, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 4, 183–190

# VERBAL STIMULI IN CONCEPT FORMATION:

## LEARNING BY DISCOVERY[1]

M. C. WITTROCK
*University of California, Los Angeles*

The 292 college students were taught by 1 of 4 treatments to decipher transposition codes. The treatments were: rule and answer given, rule given and answer not given, rule not given and answer given, and rule and answer not given. A $2 \times 2$ linear hypothesis design was used to analyze the results. It was hypothesized and verified ($p < .01$) that when retention and transfer to similar examples are the criteria, giving rules is more effective than not giving rules. An "intermediate" amount of direction produced the greatest retention and transfer, but a "maximum" amount of direction produced the greatest initial learning. The "minimum" direction group was least effective on initial learning, retention, and transfer. It was concluded that under certain conditions, practice and reinforcement for the guided development of concepts can affect retention and transfer.

The recent research on learning by discovery is reviewed by Ausubel (1961) and by Kersh and Wittrock (1962). The first point of this paper is that there is a lack of consistent, adequate terminology and labels to describe the stimuli employed in these recent studies. For example, what is described as intermediate direction in one experiment (Kittell, 1957) may be more directive than what is described as maximum direction (Craig, 1956) in another experiment. What is described as a "discovery" treatment (Gagné & Brown, 1961) may involve giving to the students the rules, the specific answers to the questions, and a second day's rehearsal over the same material that was closely directed the first day.

Within a given experiment the same labels may be used to describe stimuli, hypothetical covert behavior of the learners, and their overt responses. For example, stimuli presented to learners are sometimes described as the rote treatment. The subjects are called the rote learners. The overt responses of these learners are dubbed the rote responses. One is not surprised when in the Discussion section it is tautologically concluded that the rote learning group learned by rote.

Until the important stimulus parameters are known, the labels for the stimuli probably should be descriptive, e.g., rule given and answer given, without inferring a causal relationship between stimuli and responses. The label discovery should probably not be used to describe stimuli or treatments because it implies a relationship between stimuli and covert responses which may be unwarranted. It may also cause the reader to group together quite different treatments which should be kept separate.

The second point of this paper concerns the effects upon concept formation of the use of verbal stimuli. In research

---

[1] This research was supported in part by a University of California Faculty Research Grant.

The author wishes to thank the Western Data Processing Center for the extensive use of its IBM 7090 computer; P. Twelker of the School of Education, who prepared the necessary orthogonal equations and who supervised the processing of the data; and L. Hayward of the Division of Biostatistics, who was the program consultant.

on learning by discovery, the central question is whether or not the experimenter's presentation of verbal stimuli, e.g., rules, labels, and explanations, retards or facilitates the development of concepts. By adding a verbal stimulus to a class of stimulus objects, the members of the class may become more similar to one another and less similar to stimuli outside the class. Concept formation, as measured by transfer, may be related to the meaning attached to these verbal stimuli.[2] The recent research literature on learning by discovery has much to say about the effects of experimenter supplied verbal stimuli.

The recent literature on discovery learning is examined below for the types of stimuli which were actually given to the learners. The labels used by the researchers to describe these stimuli were viewed with caution. The principles of reinforcement, practice, and extinction were applied to the responses which the stimuli probably produced. Much apparent inconsistency in the reported results seems to disappear.

When retention of learned rules or transfer of learned rules to new examples is the criterion, giving rules appears to be more effective than either less direction (not giving rules) or more direction (giving rules and giving answers to examples). Kittell's (1957) intermediate direction group, Kersh's (1962) rote learning group, Craig's (1956) directed group, Katona's (1940, p. 36) understanding group, and Forgus and Schwartz' (1957) observer groups are all examples of giving of rules. Each of these groups was found to produce retention and transfer of learned rules as well as or better than any other group in the respective experiments. The Kersh (1958) experiment does not agree with the above studies. However, the results were explained by Kersh as due to motivation and to postexperimental practice.

In a rule given situation, the subjects are usually given simple (probably meaningful) rules and usually receive much practice and reinforcement for applying them to particular problems. This practice and reinforcement plus the distinctive verbal cues appear to enhance retention and transfer as measured on tests which contain only stimuli to which the learned labels, concepts, and overt responses lead to positive transfer. The subjects cannot discover the rule but they can discover how to apply the rule. The practice of applying the rule may also allow them to extinguish the associated but incorrect responses which they have in their repertoire.

Whenever a rule given and answer given group was included in the above experiments (Kersh, 1962, directed learning group; Kittell, 1957, maximum direction group), transfer and retention for this group was inferior to the rule given group. Haslerud and Meyers (1958), however, found no difference on a transfer test between the "specific directions given" items, and the "no directions given" items. But their experimental design may have contaminated the treatments. Each subject was given equal practice on each of the two experimental conditions. If practice on discovering concepts tends to generalize to related items, one would expect that their treatments were contaminated.

Giving the specific answers and the rules may retard retention and transfer because there is insufficient practice and reinforcement for overtly applying concepts. There is also little opportunity for strongly learned and strongly associated incorrect responses to be elicited and to undergo extinction.

If the criterion is initial learning of a limited number of specific answers rather than the transfer or retention of what has been learned, the most highly directed groups, i.e., the rule and answer given groups, do as well as or better than the other groups (Haslerud & Meyers, 1958; Kittell, 1957). Where short-term memory of a few associations is sampled, it seems that practice and reinforcement may be applied more effectively to the conditioning of these associations than to the development of concepts. Again what is practiced and reinforced appears to be what is learned.

---

[2] For an excellent discussion of labeling see Dollard and Miller (1950, Ch. 18).

Few of the above studies included a measure of the effects of the treatments upon the ability of the learners to produce and to apply alternative rules, to do divergent thinking, to learn how to discover, to create, etc. With these criteria the groups which are given practice and reinforcement for the development of these alternative and higher order rules, i.e., the discovery groups, should be the most effective. However, if the practice results in extinction rather than in regular or intermittent reinforcement, these groups may not be effective.

Although what little pertinent research there is in the literature does tend to support these statements, the support is not strong. Gagné and Brown (1961) found that their discovery and guided discovery groups were superior to their rule and example group at developing new rules. The former two groups were given practice at deriving rules but were reinforced for deriving only the correct rule.

Although giving rules and answers may provide many correct answers and thus reinforcement, the nature of this type of reinforcer must be compared with the nature of the reinforcer obtained from discovery. Independent effort is highly valued by many members of our society. By attempting several responses before producing a correct or divergent response, which is reinforced, the socially valued motive of independent effort is intermittently reinforced. Assuming a learner history of reinforcement for independent effort, persistent behavior may be produced. It would also be expected that the less directed techniques would have higher affectivity loadings (either positive or negative) than the more directed techniques.

From the above summary of research, it was predicted that when giving or not giving rules and answers to examples are compared:

1. On initial learning of specific responses, giving rules is more effective than not giving rules. Giving both answers and rules is more effective than giving either the answers or the rules. Giving neither rules nor answers is least effective.

2. On retention and/or transfer to new and similar examples, giving of rules is more effective than not giving rules. Giving of rules is more effective than giving both rules and answers or giving neither rules nor answers.

## METHOD

### *Subjects*

The subjects were 292 University of California, Los Angeles, students from two upper division educational psychology classes which were taught by the author. Participation of the subjects in the experiment was mandatory for all students enrolled in the classes and was part of the laboratory requirement of the course. There were 39 males and 253 females. Table 1 presents some information about the subjects.

### *Experimental Design*

The subjects were individually assigned at random to the eight experimental conditions. A 2 × 2 × 2 factorial, linear hypothesis design (Kempthorne, 1952, pp. 38–67) was used. The three factors were: (*a*) Rule—given or not given; (*b*) Order—example first, rule second, or rule first, example second; and (*c*) Answer—given or not given. The order factor was de-

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR THE AGE IN YEARS, VERBAL SCORE, AND NUMBER ABILITY MEASURES OF THE TREATMENT GROUPS

| Group | Age | | Cooperative English score | | Arithmetic Concepts score | |
|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Rule given and answer given ($R_gA_g$; $N = 81$) | 22.05 | 5.31 | 61.59 | 6.52 | 32.60 | 7.75 |
| Rule given and answer not given ($R_gA_{ng}$; $N = 75$) | 20.99 | 3.54 | 60.82 | 6.27 | 32.49 | 7.72 |
| Rule not given and answer given ($R_{ng}A_g$; $N = 69$) | 21.01 | 3.44 | 62.49 | 6.81 | 33.22 | 6.99 |
| Rule not given and answer not given ($R_{ng}A_{ng}$; $N = 67$) | 21.76 | 4.07 | 62.33 | 6.78 | 32.64 | 7.38 |

Note.—The Cooperative English Test, Form S (Cooperative Test Division, 1950), was used to measure verbal ability. A measure of number ability was obtained from the Arithmetic Concepts Test, a University of California, Los Angeles, product which has not been published.

signed only as a control for order of initial presentation of stimuli. The programed booklets allowed the subjects to advance to later pages or to return to earlier pages. All possible combinations of these three factors produced four experimental treatments (each with two different orders) which were designed to vary in amount of direction from the giving of examples with both rules and answers to the giving of examples without rules or answers. The four experimental treatments are listed below.

1. Rule given, answer given ($R_gA_g$)
2. Rule given, answer not given ($R_gA_{ng}$)
3. Rule not given, answer given ($R_{ng}A_g$)
4. Rule not given, answer not given ($R_{ng}A_{ng}$)

## *Materials*

Programed booklets presented the learning and initial test materials to all subjects. The booklets were of 24 weight, grey, opaque paper and were 8½ inches wide, 3¾ inches high, and 37 pages in length. The booklets for all the treatments were identical to one another in appearance. Pages 1 and 2 presented general directions on the content and use of the booklet. Pages 3, 4, and 5 presented, in various sequences depending upon the experimental treatment, appropriate directions, a rule, a worked example of the rule, and a "to be worked" example of the rule. Page 6 summarized the specific, detailed instructions for the respective experimental treatment.

The remaining 31 pages, with the exception of the last page, contained the learning material and the initial test. These 30 pages were divided into 10 series of 3 pages each. Each set of 3 pages was considered as a brief program and consisted of a page where a rule was given or was to be derived by the subject and a page where an enciphered sentence (an example of the rule) was presented followed either by the same sentence deciphered (answered) or by a space in which the sentence was to be deciphered by the subject. The order of these above 2 pages was varied to present the stimuli in one of two different orders. The subjects who were given rules, answers, or both were required to copy the respective rules and answers in spaces provided in the learning booklets. This was done in an attempt to hold constant across treatments overt practice on writing the rules and answers. The third page of each set of 3 pages was identical for all treatment groups and presented another enciphered sentence to be deciphered by the subject. This criterion sentence was also always an example of the rule which was appropriate for the example from 1 of the 2 immediately preceeding pages. These 10 criterion sentences comprised the learning test. The 10 rules, each with its two examples, were arranged in random order in the booklets. The same 20 enciphered sentences (examples) were presented to all subjects of all treatment groups.

Knowledge of correct results was not provided to the subjects. This was intentionally done to avoid contaminating across treatment groups the reinforcement and feedback which the subjects obtained when they correctly deciphered sentences.

The 3-week retention test consisted of 24 enciphered sentences to be deciphered by the subjects. The sentences were chosen to sample retention and transfer. Eight of the sentences were identical to sentences presented 3 weeks earlier. Another 8 of the sentences were "new" examples of 8 of the 10 rules presented earlier. Eight of the sentences were new examples of new transposition coding rules.

Sentences from the learning and testing materials were 18 letters long. The rules, including those of the transfer tests, were rules for deciphering transposition codes (Smith, 1943). The sentences, when properly decoded, were complete sentences and were sensible. The 24 sentences were arranged in random order. Examples of a rule, an enciphered sentence, and a correctly deciphered sentence from the learning materials appear below.

Rule—Interchange the first two and the last two letters of each group of letters with each other; then separate into words.

Enciphered sentence—

HRNTME DTGEON RKPAHE

Deciphered sentence—

MEN THRONGED THE PARK

### *Procedure*

The subjects reported to the laboratory in groups of 30 or less and were given booklets appropriate to their experimental treatments. The different treatments existed simultaneously in the same room. The subjects worked individually. They were told to read the instructions carefully and that there was no time limit for completing the exercise. In addition, they were told that they must attempt each item, and that they should immediately return the booklet after completing it.

The posttest was given 3 weeks after the learning task. Again, the subjects were allowed as much time as they required to complete the test. Each test item was scored as correct if 16 or more of the 18 letters were correct. If 15 or fewer letters were correct, no credit was given for the item.

Each subject was told at the start of the experiment that he would be informed of his own results and of the results of the experiment. A personal letter was mailed as promised.

## RESULTS

Because some subjects did not return on the appointed day for the retention test, the general linear hypothesis model (Kempthorne, 1952, pp. 38–67), a more general model than analysis of variance, was used to avoid arbitrarily eliminating subjects to equalize the cell *n* and to obtain accurate measures of the primary and interactional effects of the three variables. Homogeneity of variance and normality of distribution were assumed for each of the eight treatment groups

TABLE 2

SUMMARY OF THE *F* RATIOS FOR THE LEARNING, RETENTION, AND TRANSFER TESTS SCORES AND TIMES

| Stimulus variables | Learning test |  | 3-week retention and transfer tests |  |  |  |  |  |
|---|---|---|---|---|---|---|---|---|
|  | Learning | Time to learn | Retention 1 | Transfer to new examples 2 | Transfer to new rules and examples 3 | Combined transfer 2 + 3 | Combined retention and transfer 1 + 2 + 3 | Time to perform |
| Rule (A) | 490.74*** | 174.93*** | 19.30*** | 11.11** | 3.94* | 10.41** | 9.84** | 2.15 |
| Order (B) | .03 | .04 | .69 | 2.92 | .68 | 2.57 | .60 | .02 |
| Answer (C) | 131.77*** | 55.97*** | 2.84 | 1.16 | .15 | .98 | .25 | 4.16* |
| A × B | .38 | 9.37** | .01 | .03 | .28 | .10 | .97 | .73 |
| A × C | 157.47*** | 34.54*** | 17.57*** | 7.29** | 12.04*** | 11.36*** | 8.78** | 3.25 |
| B × C | 1.19 | .06 | 1.34 | .79 | 1.15 | 1.62 | 1.37 | .05 |
| A × B × C | 1.23 | 3.40 | 1.34 | .48 | .03 | .26 | .01 | 1.44 |

* $p \leq .05$, $df = 1/284$.
** $p \leq .01$, $df = 1/284$.
*** $p \leq .001$, $df = 1/284$.

for each criterion, except for the learning test score, after appropriate $F$ tests and graphs of frequency distributions evidenced no reason to discredit the assumptions. For the learning test, the violation of the assumptions required that the reported significance level be examined carefully. However, with an $F$ of 490.74 (see Table 2) and with the sizable increase in the mean and decrease in the variance for the rule given groups compared with the rule not given groups, there is little doubt that these results are dependable. Tables 2 and 3 present a summary of the results of the learning test and of the 3-week retention and transfer tests. Since the Order factor injected no significant amount of variance into any of the criteria, except an interactional effect for time to learn, the Order factor was ignored in the analysis of the data.

*Learning Test*

From Table 2, the Rule factor produced a statistically dependable effect upon learning ($F = 490.74$, $p < .001$). As predicted in Hypothesis 1, the groups that were given the rule correctly decoded more sentences than did the groups that were not given the rule (see Table 3).

Hypothesis 1 also stated that the experimental groups would rank from high to low, $R_gA_g$ first, $R_gA_{ng}$ or $R_{ng}A_g$ second, and $R_{ng}A_{ng}$ third. If the $R_gA_{ng}$ and $R_{ng}A_g$ groups are combined into an "intermediate direction" group (see Table 3), the predicted rank order is obtained. However, from Table 3, the $R_gA_g$ group and the $R_gA_{ng}$ group were not dependably different from each other and each was dependably better ($p < .01$) than the $R_{ng}A_g$ group, which in turn was dependably better ($t = 12.07$, $p < .01$) than the $R_{ng}A_{ng}$ group. The groups that were given the decoded examples correctly deciphered more examples than did the groups that were

TABLE 3

TREATMENT GROUP MEANS AND STANDARD DEVIATIONS FOR THE LEARNING AND 3-WEEK RETENTION AND TRANSFER TEST SCORES AND FOR THE TIME IN MINUTES TO LEARN AND TO PERFORM

| Treatment groups | Learning test | | 3-week retention and transfer tests | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Learning | Time to learn | Retention 1 | Transfer to new examples 2 | Transfer to new rules and new examples 3 | Combined transfer 2 + 3 | Combined retention and transfer 1 + 2 + 3 | Time to perform |
| Rule given and answer given ($R_gA_g$; $N = 81$) | 9.52<br>.8 | 46.64<br>13.3 | 5.33<br>1.7 | 4.16<br>1.7 | 3.27<br>1.7 | 7.43<br>2.9 | 12.74<br>4.3 | 85.11<br>20.0 |
| Rule given and answer not given ($R_gA_{ng}$; $N = 75$) | 9.73<br>.6 | 51.12<br>14.8 | 5.83<br>1.5 | 4.93<br>1.7 | 4.04<br>1.7 | 8.99<br>3.0 | 14.81<br>4.1 | 99.52<br>45.8 |
| Rule not given and answer given ($R_{ng}A_g$; $N = 69$) | 7.59<br>2.6 | 66.84<br>21.6 | 5.29<br>1.8 | 4.03<br>1.7 | 3.57<br>1.6 | 7.59<br>3.0 | 12.86<br>4.3 | 86.38<br>29.8 |
| Combined groups ($R_gA_{ng} + R_{ng}A_g$; $N = 144$) | 8.71<br>2.1 | 58.65<br>20.0 | 5.57<br>1.7 | 4.50<br>1.8 | 3.81<br>1.7 | 8.32<br>3.1 | 13.88<br>4.3 | 93.22<br>39.4 |
| Rule not given and answer not given ($R_{ng}A_{ng}$; $N = 67$) | 2.76<br>2.2 | 103.55<br>39.1 | 4.12<br>1.8 | 3.70<br>1.8 | 2.96<br>1.7 | 6.64<br>3.1 | 10.78<br>4.6 | 87.21<br>25.3 |

Note.—There were 10 items on the learning test; and on the 3-week retention and transfer tests there was a total of 24 items, with 8 items on each of the 3 subtests. Possible score on each of these tests was the respective number of items on the test. All columns, except the second and the last columns, list mean number of correct items. The second and the last columns, which are labeled Time, list time in minutes.

given the examples which were not decoded.

There was a dependable interaction between the Rule and the Answer factors ($F=157.47$, $p < .001$). When the rule was not given, giving the answer enhanced learning. When the rule was given, giving the answer did not enhance learning.

When time to learn was compared, the $R_{ng}A_{ng}$ group required significantly more time ($p < .01$) to learn than any of the other groups. The odd-even, Spearman-Brown coefficient of reliability for all experimental treatments combined was .92.

### *Retention and Transfer Test*

From Table 2, the Rule factor produced statistically significant effects on the tests of retention and transfer. As predicted also in Hypothesis 2, the $R_gA_{ng}$ group was dependably better ($t=3.01$, $p < .01$) than the $R_gA_g$ group and the $R_{ng}A_{ng}$ group ($t=5.52$; $p < .01$) on the combined test of retention and transfer. Hypothesis 2 was supported.

The $R_{ng}A_{ng}$ group had a higher retention score than learning score. The other three groups' retention scores were lower than their learning scores (there were 10 items on the learning test but only 8 on the retention test). This "postexperimental" gain for minimally directed groups has been found before (Haslerud & Meyers, 1958; Kersh, 1958).

The interaction between the Rule factor and the Answer factor discussed on the learning test above was also statistically significant on the retention and transfer tests (see Table 2). However, the Answer factor, which was statistically significant on the learning test, was not statistically significant on the retention or on either of the transfer tests. It seems that giving specific answers improved initial scores whether or not rules were given, but improved retention and transfer scores only when rules were not also given. When time to perform was considered, the groups given answers required significantly less time ($p < .05$) than did the groups not given answers.

The odd-even, Spearman-Brown reliability coefficient for all treatments combined was .65 for the retention test, .77 for the transfer tests, and .82 for the combined retention and transfer tests.

## DISCUSSION

Before the results are discussed, several limitations to their generality should be mentioned. The subjects were college students. The experimental treatments must have interacted with the subjects' strategies for learning information. The type of verbal stimuli which was studied was limited. The rules were all simple ones which did not need explanations or derivations to make them meaningful. This study did not sample ability to do divergent thinking or ability to transfer to quite different situations, e.g., to decipher substitution codes. It is probable that less directed procedures are most effective for these criteria.

The results of this study are in very close agreement with the interpretation of recent research presented in the introduction. When the criterion is initial learning of a few responses, explicit and detailed direction seems to be most effective and efficient. When the criteria are retention and transfer, some intermediate amount of direction seems to produce the best results.

If a criterion of percentage of retention were used, the results of this study would be interpreted to agree with those of Haslerud and Meyers (1958) who did use that criterion. Far different conclusions from those given above would be presented because only the $R_{ng}A_{ng}$ group produced an increase in score from the learning test to the retention test. It is extremely important that one relate his results to the criterion which was sampled.

The study implies that under certain conditions meaningful direction can supply distinctive verbal cues which can increase the probability that the learner responds in the directed fashion and, therefore, can decrease the probability that he responds in other fashions. This verbal direction can be used to enhance transfer by eliminating alternative responses which produce negative transfer and by including alternative responses which produce positive transfer. It appears that the direction should match the criteria. If the direction provides explicit practice and reinforcement for responses across the range of positive stimuli and allows the responses outside this range to be elicited and to be extinguished, positive transfer may be enhanced. To refrain from introducing labels for concepts when learners have adequate backgrounds to make these labels meaningful seems to be a vain attempt to enhance transfer which instead probably only reduces it and wastes time during acquisition.

Meaningful terminology and labels, in addition to the important but general term "reinforcing stimuli," are much needed to describe and to classify the stimulus variables employed in research on discovery learning. A sophisticated theory of concept formation is needed not only to supply these terms and labels but also to account for the phenomena observed in research on learning by discovery.

## REFERENCES

Ausubel, D. P. Learning by discovery: Rationale and mystique. *Bull. Nat. Ass. Second. Sch. Principals,* 1961, **45**, 18–58.

Cooperative Test Division. *Cooperative English Test: Form S.* Princeton: Educational Testing Service, 1950.

Craig, R. C. Directed versus independent discovery of established relations. *J. educ. Psychol.,* 1956, **47**, 223–234.

Dollard, J., & Miller, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

Forgus, R. H., & Schwartz, R. J. Efficient retention and transfer as affected by learning method. *J. Psychol.,* 1957, **43**, 135–139.

Gagné, R. M., & Brown, L. T. Some factors in the programing of conceptual learning. *J. exp. Psychol.,* 1961, **62**, 313–321.

Haslerud, G. M., & Meyers, Shirley. The transfer value of given and individually derived principles. *J. educ. Psychol.,* 1958, **49**, 293–298.

Katona, G. *Organizing and memorizing.* New York: Columbia Univer. Press, 1940.

Kempthorne, O. *The design and analysis of experiments.* New York: Wiley, 1952.

Kersh, B. Y. The adequacy of "meaning" as an explanation for the superiority of learning by independent discovery. *J. educ. Psychol.,* 1958, **49**, 282–292.

Kersh, B. Y. The motivating effect of learning by discovery. *J. educ. Psychol.,* 1962, **53**, 65–71.

Kersh, B. Y., & Wittrock, M. C. Learning by discovery: An interpretation of recent research. *J. teacher Educ.,* 1962, **13**, 461–468.

Kittell, J. E. An experimental study of the effect of external direction during learning on transfer and retention of principles. *J. educ. Psychol.,* 1957, **48**, 391–405.

Smith, L. D. *Cryptography: The science of secret writing.* New York: Norton, 1943.

(Received August 21, 1962)

*Journal of Educational Psychology*
1963, Vol. 54, No. 4, 191–196

# CONCEPTUAL DISCOVERY AND CONTROL ON A PICTORIAL PART-WHOLE TASK AS A FUNCTION OF AGE, INTELLIGENCE, AND LANGUAGE[1]

HANS G. FURTH
*Catholic University of America*

In order to observe nonverbal conceptual behavior in relation to age, intelligence, and language, normative data for hearing children and deaf children, aged 6–10 and 14, were collected. The major requirement of the task was a consistent control of choice behavior by the discovered principle of part-whole. In spite of moderate correlations between IQ and performance, improvement with age was only gradual. Just as lack of verbal experience did not handicap deaf Ss, the kind of verbalization after the task was found to be uncorrelated with performance. On the basis of these findings a construct of conceptual control was postulated as being related to IQ but only minimally associated with age or language.

This study investigated children's ability to employ the part-whole concept in a nonverbal situation. By part-whole is meant the relationship of two visual objects, such that one object can commonly be called part of the other. Thus button is considered part of a coat while hat and coat, although related to each other, do not exemplify a part-whole relation. Two aspects of mastery of the concept were observed: discovery of the principle and consistent choice behavior according to the discovered principle.

The experiment was designed to shed light on the relation between nonverbal conceptual behavior and three important variables: age, intelligence, and verbal language. Normative data were collected over an age span from 6 to 14 years, and the performance of children with high and low intelligence within and between each age group was analyzed. In this way it was possible to evaluate the construct of IQ and MA in relation to this nonverbal task. Verbal language in this context refers to the ordinary language of one's society, i.e., spoken English, as part and parcel of a growing child's environment and as possibly being employed internally during the solution of a task. The influence of language was manipulated by comparing deaf and hearing children at each age level and by asking hearing children to verbalize the principle by which they operated.

Apart from clarifying the influence of verbal experience on cognitive behavior, a comparison of deaf and hearing children contributes towards the understanding of the cognitive functioning of deaf children. General vague statements about deaf persons' deficits in abstract ability have been only recently subjected to scientific inquiry (Furth, 1961, 1963; Kates, Kates, Michael, & Walsh, 1961). Many investigators have tacitly assumed that verbal language is an important or indispensable ingredient of general cognitive development and that the inferior performance of deaf people on various tasks is a direct reflection of their language deficiency. In such cases it would appear scientifically more appropriate to analyze the task variables in relation to the deaf child's total experience rather than to make unwarranted claims about the latter's deficiency in abstract thinking (Furth, in press; Milgram & Furth, 1962). From a preliminary analysis of the present task it was expected that no differences would emerge in the performance of deaf and hearing children.

---

[1] This study was supported in part by Grant 704 from the United States Vocational Rehabilitation Administration. Thanks are expressed to Sarita Berry who collected normative data. The cooperation of the St. Anthony's Parochial School in Washington, D. C., and of the schools for the deaf mentioned in the text is gratefully appreciated.

A part of this paper was read at the annual meeting of the American Psychological Association, New York, 1961.

## METHOD

### *Subjects*

The hearing children came from Grades 1–5 and Grade 10 of a parochial school system. They were divided into six groups on the basis of age, 6–10 and 14 years. The number of children in each age group was 36, 51, 87, 59, 62, and 100, respectively. The deaf children were taken from the Maryland State School, a Baltimore day school, and the New York School for the Deaf. The respective number for each age group was 16, 11, 14, 20, 13, and 41. In all groups the sexes were about evenly divided.

### *Task*

Twenty sets of five pictures each were drawn with black pencil on 3 × 5 inch white cards by a commercial artist. In each set two drawings illustrated the part-whole relationship and their choice was considered "correct." Several other reasonable combinations according to class, function, or experiential association were present on each set. These served as controls for the specificity of the attained principle by providing the pull of strong misleading perceptual associations. Table 1 presents the series of pictures shown with the "whole-part" pair at the beginning of each set.

### *Procedure*

Each child was seen individually in a small private room. The experimenter placed the five pictures of Set 1 on the table before the subject in a random order so as to discourage formation of position habits. Without additional explanation the child was told by words or gestures to pick up or point to two of the five pictures. During Learning Trials 1–13 a correct choice was reinforced by the experimenter's nodding or saying "good." If the child failed, the experimenter said "no" and pointed to the correct pair. On Trials 14–20 no correction or reinforcement was given. After

TABLE 1

STIMULUS SERIES

| Set | Picture | | | | |
|---|---|---|---|---|---|
| 1 | Pot | Lid | Basket | Cap | Plate |
| 2 | Bicycle | Wheel | Shoes | Bus | Table |
| 3 | Flashlight | Bulb | Umbrella | Fire | Ball |
| 4 | House | Chimney | Tree | Jar | Tent |
| 5 | Tree | Leaf | Boy | Flower | Cow |
| 6 | Church | Steeple | Man | House | Factory |
| 7 | Fireman | Firehat | Police | Ambulance | Fire |
| 8 | Tricycle | Handlebars | Girl | Car | Wagon |
| 9 | Letter | Stamp | Mailbox | Book | Pencil |
| 10 | Elephant | Tusk | Fish | Dog | Bowl |
| 11 | Bird | Wing | Fish | Balloon | Airplane |
| 12 | Belt | Buckle | Sweater | Rope | Pants |
| 13 | Lamp | Bulb | Tree | Fire | Sun |
| 14 | Coat | Button | Man | Flower | Hat |
| 15 | Clock | Clockhands | Calendar | Watch | Telephone |
| 16 | Babybottle | Nipple | Cow | Baby | Cradle |
| 17 | Man | Hand | Cigar | Bed | Pencil |
| 18 | Face | Ear | Eyeglasses | Book | Hat |
| 19 | Shoe | Shoelace | Feet | Slippers | Sock |
| 20 | Book | Page | Pencil | Briefcase | Pen |

TABLE 2

Mean Error of Groups as a Function of Age

| Set | Group | Age level 6 | 7 | 8 | 9 | 10 | 14 |
|---|---|---|---|---|---|---|---|
| Learning (2–13) | Hearing | 3.25 | 2.02 | 2.32 | 1.64 | 1.24 | .91 |
| | Deaf | 3.87 | 2.36 | 2.64 | 1.55 | 1.62 | .83 |
| Criterion (14–20) | Hearing | 3.92 | 4.12 | 3.45 | 3.44 | 2.42 | 2.21 |
| | Deaf | 4.12 | 4.00 | 3.64 | 3.00 | 2.92 | 2.09 |
| Total (2–20) | Hearing | 7.17 | 6.14 | 5.77 | 5.08 | 3.58 | 3.12 |
| | Deaf | 8.00 | 6.36 | 6.28 | 4.55 | 4.54 | 2.93 |

the last trial the hearing subject was asked to verbalize, if possible, the principle of his choice behavior. On a separate day afterwards all hearing subjects were given the appropriate form of the Otis Intelligence Group Test. Since results on the Otis test for first graders and for some others were unsatisfactory, the number for the part-whole task and IQ is not equal in all cases.

## Results

It seemed appropriate to separate errors on the learning trials from errors on criterion trials as well as to analyze total errors. Performance on Set 1 was disregarded as being entirely due to guessing. Table 2 summarizes average error for the hearing and deaf groups at six age levels.

A breakdown by sex showed no appreciable differences, hence the sex variable is disregarded in the analysis of data. First the performance of the hearing groups was subjected to an analysis of variance and for this purpose 36 subjects were randomly selected from each age group. The mean error score of these groups differed only a little from the one based on all subjects as given in Table 2.

Highly significant *F*s emerged for age variance with learning (9.59), criterion (8.30), and total score (12.09). Duncan's multiple-range test (Li, 1957, p. 238) for learning trials showed that the 6-year group differed from all others and that Groups 7–10 and Groups 9–14 did not differ reliably. For the criterion trials Duncan's test revealed a very gradual decrease of errors with age by constant overlapping of three age levels except for the 14-year group which only overlapped with the 10-year group. For total error score the 14-year group was found to differ significantly from all other ages, while the 6- and 7-year groups and the 8–10 year groups did not differ reliably from each other. As an overall picture then there was a gradual significant improvement of part-whole performance with age; on the learning part a marked decrease of errors took place at age 7, and for the criterion part a similar improvement at age 10.

The performance of the deaf children as depicted in Table 2 was notably similar to that of their hearing controls. Not 1 of the 18 possible comparisons approached significance by a *t* test.

To evaluate the performance of above and below average subjects, 10 subjects each were randomly selected at five age groups from all subjects with IQs between 105 and 115 and with IQs between 85 and 95. Table 3 gives the pertinent data which are in essential agreement with the ones reported for

TABLE 3

MEAN ERROR AS A FUNCTION OF AGE AND IQ

| Set | IQ | Age | | | | |
|---|---|---|---|---|---|---|
| | | 7 | 8 | 9 | 10 | 14 |
| Learning (2–13) | High | 1.2 | .8 | 1.2 | 1.4 | .5 |
| | Low | 3.8 | 3.8 | 1.9 | 1.3 | 1.5 |
| Criterion (14–20) | High | 3.5 | 2.2 | 2.7 | 1.4 | 1.9 |
| | Low | 5.3 | 4.5 | 4.0 | 3.2 | 3.0 |
| Total (2–20) | High | 4.7 | 3.0 | 3.9 | 2.8 | 2.4 |
| | Low | 9.1 | 8.3 | 5.8 | 4.5 | 4.5 |

the total sample. Learning, criterion, and total errors were analyzed and consistently the greater part of the variance had its source in IQ, with *F* ratios exceeding 20 in all cases. In contrast the overall *F* for age, although significant beyond .01, was only 4.51 for learning, 5.74 for criterion, and 6.65 for total errors. Apparently then there was a greater degree of association between IQ and error than between age and error on the part-whole task. Such a conclusion is substantiated by correlations between IQ and criterion, based on all the subjects of an age group which were .33, .51, .29, .39, and .43 for ages 7, 8, 9, 10, and 14, respectively. These moderate correlations reported differ significantly from zero.

The verbalizations of the older group only were analyzed since the great majority of even the fifth graders merely said "I can't tell." Of 115 subjects 14 and 15 years old, 22 gave the latter answer; 46 said something like "go together, belong"; 30 mentioned "needed" or "repeated"; and only 17 used the correct verbal description "part." Table 3 shows the number of errors of these four groups on the seven criterion trials. The order of index association (*o*; Senders, 1958, p. 130) between verbalization and performance is a weak, but significant *o* of .15.

Finally, types of errors were analyzed to discover possible age-characteristic choices. Thus on Set 17 an incorrect choice of man and coat or on Set 19 a choice of a sock and feet was made frequently but 14- and 8-year-old subjects erred with about similar frequency. In general, then, all age groups and hearing and deaf subjects were subject to similar kinds of errors.

## DISCUSSION

While it is not surprising that performance on the task improved with age, the slowness of the decrease in errors is quite striking. Fourteen-year-old children failed to act in a logically consistent way nor have they reached the stage where they can be said to attack the present task in a qualitatively different way from much younger children.

TABLE 4

FREQUENCY OF ERRORS ON SEVEN CRITERION TRIALS AS RELATED TO TYPE OF VERBALIZATION

| Verbalization | Errors | | | | | |
|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 |
| Part | 5 | 5 | 5 | 2 | 0 | 0 |
| Needed | 7 | 4 | 6 | 3 | 4 | 6 |
| Go together | 5 | 3 | 10 | 16 | 9 | 3 |
| Can't tell | 2 | 6 | 6 | 5 | 1 | 2 |

It would appear that the major requirement of the task, besides discovering the principle of choice, was to adhere to the concept in spite of competing pulls in other misleading directions. In line with previously reported results on a verbal task (Schooley & Hartman, 1937) even the younger children had little difficulty in discovering the principle and relatively few errors were made during learning trials.

Conceptual control would seem an appropriate term for describing the nature of the present task. Luria (1961) spoke of verbal control of behavior and the above term is purposely related to Luria's term. Luria's procedure, however, employed verbal instructions and he reached his conclusion by observing the influence of these instructions on behavior. The present procedure was nonverbal and control of behavior seemed to be largely independent of verbal behavior. Such an inference appears justified whether one thinks of language in terms of general verbal experience or as specifically facilitating and controlling a particular task. Deaf children are certainly minimally exposed to verbal experience and were found to perform very much like their hearing peers. Hearing children, on the other hand, verbalized very little, and where they did so, the relation of verbalization to performance was tenuous at best. Many subjects verbalized wrongly and succeeded fairly consistently and, in contrast, of the 17 subjects who verbalized correctly, only 5 made consistently correct choices on Criterion Trials 14–20. It would be instructive to observe more directly the influence of verbal control by asking children verbally to base their choice on the part-whole concept.

The curious fact of a moderate correlation of part-whole performance with IQ, yet not with age, deserves some attention. A look at Table 2 will show that the criterion score for ages 6–9 inclusive differed but little, yet correlations ranging from .29 to .51 were reported between error and IQ. At each age level as depicted in Table 3 the lower IQ group made significantly more errors than the higher group. A comparison of the rather similar performance of the high 7-year and the low 14-year groups illustrates that our traditional MA is an inappropriate concept in connection with the present task. The average MA of the lower 14-year group was about 12-7 years with a mean error of 4.5 on the part-whole task. According to Table 2 a total error score of 4.5 corresponds to an age level of 9 or 10 which is 3 years below the MA level of the below average 14-year subjects.

While such points show that the present task would be unsuitable for a traditional kind of ability test, it does seem to highlight a difference between below and above average mental ability which may have much to do with what is here called conceptual control. The more intelligent child may be the one who, regardless of relatively great age differences, can control his behavior more consistently and more successfully by means of conceptual principles.

The remarkable similarity of deaf and hearing groups' performance demonstrated that deaf children were not inferior in discovery of the part-whole concept nor in conceptually controlled behavior. Both findings appear contrary to older generalizations but are well in accord with more recent findings (Kates, Yudin, & Tiffany, 1962). Since none of the three public schools from which deaf subjects were drawn pursue a selective policy with respect to admission, there is no ground for the assumption that the deaf subjects represent an above average sample. While it is readily admitted that the concept of part-whole by itself is of no particularly high order of abstraction, concept controlled behavior, however, would appear to be part of what Goldstein and Scheerer (1941) called the abstract attitude.

## REFERENCES

FURTH, H. G. The influence of language on the development of concept formation in deaf children. *J. abnorm. soc. Psychol.*, 1961, **63**, 386–389.

FURTH, H. G. Classification transfer with disjunctive concepts as a function of verbal training and set. *J. Psychol.*, 1963, **55**, 477–485.

FURTH, H. G. Conservation of weight in deaf and hearing children. *Child Develpm.*, in press.

GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior: An experimental study with special tests. *Psychol. Monogr.*, 1941, **53**(2, Whole No. 239).

KATES, S. L., KATES, W. W., MICHAEL, J., & WALSH, T. M. Categorization and related verbalizations in deaf and hearing adolescents. *J. educ. Psychol.*, 1961, **52**, 188–194.

KATES, S. L., YUDIN, L., & TIFFANY, R. K. Concept attainment by deaf and hearing adolescents. *J. educ. Psychol.*, 1962, **53**, 119–126.

LI, J. C. R. *Introduction to statistical inference.* Ann Arbor, Michigan: Edwards Brothers, 1957.

LURIA, A. R. *The role of speech in the regulation of normal and abnormal behaviour.* New York: Pergamon Press, 1961.

MILGRAM, N. A., & FURTH, H. G. Conceptual categorization and verbalization to pictorial and verbal stimuli in normal and educable retarded children. *Amer. Psychologist*, 1962, **17**, 303–304. (Abstract)

SCHOOLEY, M., & HARTMAN, G. W. Role of insight in the learning of logical relationships. *Amer. J. Psychol.*, 1937, **49**, 287–292.

SENDERS, VIRGINIA L. *Measurement and statistics.* New York: Oxford Press, 1958.

(Received November 19, 1962)

*Journal of Educational Psychology*
1963, Vol. 54, No. 4, 197–201

# IMPLICIT REINFORCEMENT OF RESPONSES[1]

LEE SECHREST
*Northwestern University*

The question under examination was the effect on children working in pairs of either positive or negative reinforcements administered to their pairmates. 2 groups of 15 pairs of children each were run in an experiment involving time taken to solve 2 puzzles. Between the 2 puzzles 1 randomly chosen child was either positively or negatively verbally reinforced. A control group of 15 pairs was also run without reinforcement. Whether reinforcement was explicit or implicit made no difference, but it was found that positive reinforcement facilitated and negative reinforcement hindered performance. The group that was implicitly reinforced by observing their pairmates subjected to reproof was significantly better than all other groups. Implicit reinforcement effects are thought to be limited to small groups and competitive situations.

There are many ways in which an organism, particularly a human, may gain information about the outcomes of responses which he makes or can make. The most obvious way is by making a response and then either enjoying or suffering its consequences. Such is what is meant by direct, explicit reinforcement. For the most part psychologists have devoted themselves to the study of the effect on an indiviual's behavioral dispositions of the reinforcements directly contingent upon his own responses. However, among the other ways in which response consequences may become evident is by the observation of the experiences of other persons (Campbell, 1961, 1963). Under some circumstances an individual may observe another person behaving, observe reinforcements being delivered to the other person in such a way that they seem contingent upon particular responses, and thereby have his own response tendencies altered in a congruent manner. Much of what has been called imitation or observational learning (Miller & Dollard, 1941; Rosenbaum, 1961) is of that general nature. Berger (1961), using the term "vicarious reinforcement" suggested by Lewis and Duncan (1958), has been able to demonstrate the learning of words and numbers on an incidental basis as a function of reinforcements delivered to a participant. More recently he has been able to demonstrate a conditioned GSR in subjects who were watching someone else receive what they thought was an electric shock (Berger, 1962).

On the other hand it is possible that under different circumstances an observer might watch the reinforcements being delivered to a model and have his own behavior tendencies altered in a direction opposite to the alterations produced for the model. Thus, for example, Briggs (1927, 1928) reported the retrospective accounts of early school experiences by high school and college students. He found that his respondents rather consistently reported having been adversely affected by sarcasm directed at them by a teacher. However, they also rather consistently reported that observing another pupil become the object of a teacher's sarcasm had a salutary affect on their own behavior. Thus the apparent effects produced for the observer and for the actual recipient of the reinforcement are opposites. If it is possible for an observer to make some inference about the adequacy of his own responses from the observations of reinforcements delivered to a model, then the observer may have his own response tendencies altered and we may say that he has been implicitly reinforced. For example, if a mother comes upon two of her children drawing a picture and says to one of them, "My that's a very nice picture you're drawing," and says nothing to the second child, it would not be surprising if the second child supposed that his own picture was by comparison inferior and if he then behaved as if he had been negatively rather than positively reinforced. In such a case we might say that the observing child was implicitly reinforced. It has been found that children are quite aware of verbal reinforcements administered to their classmates (Sechrest, 1962).

---

[1] This investigation has been supported by Northwestern University's psychology-education project, sponsored by the Carnegie Corporation. The writer wishes to thank Nancy KcKiernan for assistance in data collection and J. Wallace for a careful reading of the manuscript.

One type of situation in which implicit reinforcement of an observer is quite likely to occur, i.e., in which the effect produced on the observer is the opposite of that produced on the model, is one analogous to a zero-sum game, or in which the observer holds to a "depleted pool" hypothesis about available reinforcements. If a reinforcing agent has only so many reinforcements to dispense, then every reinforcement delivered to one person reduces the possibilities of reinforcement for every other person. For example, if a parent has only one piece of candy with which to reward two children, the dispensing of the candy to one child would not necessarily be expected to produce a congruent effect in an observing child.

In this paper an experiment is reported in which young school children were engaged in pairs in an experimental task during which the experimenter made a comment to one of the children which apparently made it possible for the other children to make some inference about the adequacy of his own performance and thus implicitly rather than explicitly be reinforced. For the sake of argument it is assumed that implicit positive reinforcement occurs in the pair in which explicit negative reinforcement is given. Similarly implicit negative reinforcement should occur when direct positive reinforcement is given.

## PROCEDURE

Preliminary work showed that it was possible to affect the speed with which children would work on a puzzle task by employing general verbal reinforcements. For this study the experimental task was the completion of a pair of wooden jigsaw puzzles of the type usually found in toy stores. The puzzles were of a moderate level of difficulty, having 18 and 21 pieces, respectively. The data for analysis were the difference scores between the log time scores (log transformation to correct for skewness) for the first and the second of the two puzzles.

In order for implicit reinforcement to occur the subjects must have the opportunity to make observing responses; therefore, children were run in pairs. Each child of a pair was given one of the two puzzles to work, and following completion by both children of their puzzles, each child was given the second one, i.e., they exchanged. Thus, all the subjects worked the same two puzzles, but in differing orders. The experimental manipulation occurred in the interval between the two puzzles. In the Explicit Reinforcement group one of the two children, randomly selected, was told,

> My, you did very well with your puzzle. That was good work. Now we are going to work another one.

In the Explicit Negative Reinforcement group one child of each pair, again randomly selected, was told,

> Well, that really wasn't so good. You seemed to have trouble with that one. Now we are going to work another one.

And finally in a control group the experimenter merely said,

> All right, now we are going to work another one.

Thus there were two experimental groups in which one of the children received either a positive or a negative verbal reinforcement and a control condition in which no evaluative comments were made. Time scores were recorded to the nearest 5 seconds, and surreptitiously, so that the children were not aware that speed was an issue.

Fifteen pairs of children were run in each experimental group, making 90 children in all. Children were selected in pairs by their teacher as being at about the same ability level, and all pairs were same-sex pairs. The children came from kindergarten, first, second, and third grade classrooms in the three schools in nearly equal proportions for the three reinforcement groups. Experimental procedures were conducted by a young female graduate student in education. The children were asked to go with the experimenter "to play some games." Every attempt was made to keep overt competition between the children at a minimum by using children at the same ability level, by having them work on different puzzles, by structuring the situation as a game session, etc. It will be noted that the verbal reinforcements used were quite general in nature and did not refer at all to time scores which were the measure of performance.

In the positive and negative reinforcement groups the children could be divided into those who had received explicit and those who had received implicit reinforcement. There were 15 of each. The 30 children in the control group were, of course, undifferentiated.

## RESULTS

The first analysis was a 2 × 2 analysis of variance for the two types (positive and negative) and the two conditions (explicit and implicit) of reinforcement, the data for which are given in Table 1. The results of the analysis are presented in Table 2. It can be seen that whether the reinforcement was explicit or implicit was of no consequence, but that positive and negative verbal reinforcements produced a significantly different effect, with positive reinforcement tending to have a facilitating effect on speed of performance on the second puzzle. There is no significant interaction between type and condition of reinforcement.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF THE DIFFERENCE BETWEEN LOG TIME OF TWO PUZZLES FOR FOUR EXPERIMENTAL GROUPS AND ONE CONTROL GROUP

| Group | *N* | *M* | *SD* |
|---|---|---|---|
| Explicit positive | 15 | −.46 | 1.81 |
| Implicit negative | 15 | .27 | 1.88 |
| Explicit negative | 15 | .19 | 1.80 |
| Implicit positive | 15 | −1.43 | 1.56 |
| Control | 30 | − .12 | 2.11 |

A comparison of the four experimental groups with the control group shows that only the implicit-positive group differs significantly from the controls ($t = 3.20$, $p < .01$). The explicit-positive and implicit-positive groups also differ from each other ($t = 2.05$, $p < .05$), although this finding is of questionable meaning in view of the lack of a significant main effect for condition of reinforcement.

## DISCUSSION

We have seen that a negative reinforcement delivered to one child apparently operates as a positive reinforcement for another child who is par-

TABLE 2

ANALYSIS OF VARIANCE OF LOG TIME SCORES FOR FOUR EXPERIMENTAL GROUPS AND ONE CONTROL GROUP

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Type of reinforcement (positive or negative) | 1 | 21.14 | 6.32* |
| Condition of reinforcement (explicit or implicit) | 1 | 2.90 | < 1.00 |
| Type × condition | 1 | 4.07 | 1.22 |
| Within groups | 56 | 3.34 | |
| Total | 59 | | |

* $p < .02$.

ticipating in the same activity. Conversely, positive reinforcement intended explicitly for one child may have an implicitly negative effect on another. In fact in the experiment just described there was no overall difference attributable to the condition of administration of the reinforcement, i.e., whether it was explicit or implicit. The principal result in the experiment was the difference between positive and negative reinforcements in producing an improvement or a decrement in the speed of solution of the puzzles. However, the more detailed consideration of the results provided by *t* tests suggests that the implicit positive reinforcement was particularly likely to affect performance on the second puzzle. It will be remembered that none of the other three groups differed significantly from the control group, which showed a slight tendency toward improvement in time scores from Puzzle 1 to Puzzle 2. It is of interest to note that the results are somewhat reminiscent of the responses obtained by Briggs (1927, 1928), for his subjects reported being positively affected by negative verbal reinforcement administered to other persons. They also did not report being as susceptible to positive comments made to other persons, i.e., comments which in a competitive situation would be seen as implicitly negative. It should also be pointed out that Kounin and Gump (1958) and Gnagey (1960) have reported experimental findings suggesting the powerful effects on children of observing other children being subjected to a teacher's discipline.

We may ask what the conditions are under which subjects may act as if they had been vicariously reinforced, i.e., will act as if they themselves had received explicit reinforcement, and the conditions under which they will be expected to behave as if they had received an implicit, that is, an opposite, reinforcement. First of all it seems probable that subjects will behave as if they had received an implicit reinforcement when comparisons between performance are fairly evident. When the evaluative dimension of performance is not relevant, i.e., when there is no expectation on the part of the subject that his performance is to be evaluated, then there would seem to be no particular reason why he should respond to a positive reinforcement to a model as if he himself had received a negative reinforcement. It is to be noted in Berger's (1962) experiment as well as in some of the experiments of Rosenbaum (1961) and the experiment of Lewis and Duncan (1958) that the experimental procedures did not involve an evaluation of the responses of either the model or the observer. Hence, there was no reason for the subject to question the adequacy of his own performance.

Secondly, it would seem that implicit reinforcing effects would be more likely to be found in situations in which the observer and the model are engaged in some sort of competitive or quasi-competitive task. If the observer and the model are working on very different kinds of tasks, then an explicit reinforcement to the model would seem to have fewer implications for the adequacy of the performance of the observer. If one child is drawing a picture and the second child is jumping rope and an adult comes along and says, "My that's very nice," to the child drawing a picture, the child jumping the rope might well not be affected in any way, at least in terms of his rope jumping. He might, however, decide that he would rather draw pictures than jump a rope. Third, implicit reinforcing effects would probably be minimumized in groups of a large size. If in a classroom of 30 children a teacher tells 1 child that his paper is very good, the implicit invidious comparison of other children's papers is probably minimal.

## REFERENCES

BERGER, S. M. Incidental learning through vicarious reinforcement. *Psychol. Rep.*, 1961, **9**, 477–491.

BERGER, S. M. Conditioning through vicarious investigation. *Psychol. Rev.*, 1962, **62**, 450–466.

BRIGGS, T. H. Praise and censure as incentives. *Sch. Soc.*, 1927, **26**, 596–598.

BRIGGS, T. H. Sarcasm. *Sch. Rev.*, 1928, **36**, 685–695.

CAMPBELL, D. T. Conformity in psychology's theories of acquired behavioral dispositions. In I. A. Berg & B. M. Bass (Eds.), *Conformity and deviation.* New York: Harper, 1961.

CAMPBELL, D. T. Social attitudes and other acquired behavioral dispositions. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 6. *Investigations of man as socius: Their place in psychology and the social sciences.* New York: McGraw-Hill, 1963. Pp. 94–172.

GNAGEY, W. J. Effects of classmates of a deviant student's power and response to a teacher-exerted control technique. *J. educ. Psychol.*, 1960, **51**, 1–8.

KOUIN, J. S., & GUMP, P. V. The ripple effect in discipline. *Elem. Sch. J.*, 1958, 158–162.

LEWIS, D. J., & DUNCAN, C. P. Vicarious experience and partial reinforcement. *J. abnorm. soc. Psychol.*, 1958, **57**, 321–326.

MILLER, N. E., & DOLLARD, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

ROSENBAUM, M. E. Imitation and observational learning. Colloquim given at Northwestern University, 1961.

SECHREST, L. The motivations of young children in school: Some interview data. *J. exp. Educ.*, 1962, **30**, 327–335.

(Received November 26, 1962)

*Journal of Educational Psychology*
1963, Vol. 54, No. 4, 202–207

# EXPRESSION OF CONCEPTS THROUGH WRITING AND DRAWING AND EFFECTS OF SHIFTING MEDIUM

NANCY ECHOLS WILDER[1] AND DONALD ROSS GREEN
*Emory University*

The frequently noted difficulty in verbalizing apparently learned concepts suggested that shifting expression of a concept from writing to drawing or drawing to writing would interfere with expression. 88 4th graders observed 2 plants growing for 2 weeks—1 in light, 1 in shade. They then described the differences (giving discrimination scores) and explained the reason for the differences (giving cause-effect, or CE, scores). 2 similarly treated but unobserved plants were shown next and another expression obtained. Ss either wrote twice, wrote then drew, drew then wrote, or drew twice. It proved easier to draw differences and easier to explain in writing. Shifting medium did not influence discrimination scores. However, CE scores were interpreted as favoring the hypothesis.

If explicit verbal rules are given, subjects often show less ability to apply or transfer these rules than those in groups asked to develop the concepts for themselves (e.g., Katona, 1940; Kersh, 1958; Szekely, 1950). Similarly verbalizing inductively derived principles and concepts is often reported to be difficult (e.g., Buswell & Kersh, 1956; Corman, 1957; Smoke, 1932). Further it has been suggested that when subjects required to verbalize such derived concepts have difficulty in doing so the ability to apply or use them is inhibited (Hendrix, 1948; Schwartz, 1955).

A noteworthy fact about all these studies is that a shift from one means of expressing the learning to another was involved. In each, either pictorial, mathematical, or some other form of symbolic representation of the ideas was used initially and verbal expression followed, or vice versa. Thus, taking the use of a principle as a form of expressing the idea, both kinds of difficulty noted above can be explained as follows: a shift from the expression of an idea in one medium to expression in another medium tends to create difficulty and confusion. A subject who derives an arithmetic principle for himself can use it more readily than one who is told about it in words because no shift in form of expression is involved, but he will have more difficulty in describing the principle verbally because such a shift is involved.

It may be noted that different media of expression often require the use of different sets of symbols or symbol systems which, for convenience, we will call codes. It seems probable that the difficulty in shifting medium, hypothesized above, is at least partially because of the shift of code involved.

The purpose of this study was to explore some aspects of these ideas. Two media of expression involving two very different codes, drawing and writing, were used by children to express an idea. The children observed a demonstration and then were asked to express the concept demonstrated by writing or by drawing a picture. A second expression of the concept was then obtained. For half of the children, the second expression was in the same code as the first and for half the children the second expression was in a different code from the first.

---

[1] Now at Westminster Schools.

Differences in the adequacy of expression in the two codes were expected. Drawing is more representational of the actual appearance of objects than is writing. Writing uses languages and is a more customary way of communicating. Thus, it was anticipated that on the average the two codes would prove approximately equal in difficulty.

It was hypothesized that those subjects required to shift codes from the first expression to the second expression would show more loss or less gain on the second expression than those using the same code both times.

## METHOD

### *Subjects*

The subjects were 88 fourth grade students drawn from four classes in four rather diverse schools in a surburban community.[2] The system uses a sort of three-track program and all subjects were drawn from those placed in "advanced studies" because a pilot study had indicated that the task was a bit difficult for subjects this age. The mean IQ of the group was approximately 115 on the California Test of Mental Maturity (CTMM).

### *Procedure*

The experimenter told each class that this was an experiment and they were asked to watch closely and to listen carefully. The experimenter then made two very similar cuttings from a coleus plant, put them in clear glass jars, and placed one in a sunny window and the other in poor light. Subjects were asked to observe the plants during the next 2 weeks but not to talk to anyone about the plants during that time.

Exactly 2 weeks later, the experimenter returned and assigned the children randomly to one of the four experimental treatments. There were four groups in each classroom, each yielding two expressions of the concept as shown in Table 1.

Paper, mimeographed instructions, and crayons of appropriate colors were distributed. Then the two coleus plants were placed together where all could see them equally well. Identical instructions were given to all subjects except that half were instructed to write and half to draw. The children were asked to indicate the specific differences between the plants and indicate why they were different. No time limits were set. Expression I was completed by all subjects in each class after 45 minutes. Almost all subjects had a chance to relax and read before beginning Expression II.

Then the coleus plants were removed and replaced by two begonia plants which had not been seen previously. Each subject was then assigned to either write or draw according to the prearranged schedule. Instructions and procedure were the same as for the preceding expression except that the subjects were told that these begonia plants had been grown under the same conditions as the coleus plants. Everybody was finished in about 30 minutes this time. Drawing typically took longer than writing on both expressions.

### *Scores*

Twelve adults served as judges. Each student's work was assigned 0, 1, or 2 points for indications about why the plants differed. This will be called the cause-effect (CE) score. Both plant pairs differed in five major ways: root length, number of leaves, leaf size, leaf color, amount of leaf wilt. In addition, one of the two begonias used for Expression II was blooming. One point was given for each of the differences indicated by a subject yielding

TABLE I

EXPERIMENTAL DESIGN

| Group label | Code used on Expression I | Code used on Expression II | Number of subjects |
|---|---|---|---|
| WW | Write | Write | 22 |
| WD | Write | Draw | 22 |
| DW | Draw | Write | 22 |
| DD | Draw | Draw | 22 |

---

[2] Originally three other classes in three additional schools were included. Each of these classes inadvertently received instruction relevant to the task during the experiment. CE scores of these classes differed from those reported here. Since instruction varied from class to class no meaningful interpretation of these differences could be devised.

a discrimination score. Thus on Expression I a discrimination score of 5 and a total score of 7 was possible, while on Expression II 6 and 8 were possible. Expression I discrimination scores were multiplied by 6 and Expression II discriminations by 5 to adjust for this discrepancy in possible discriminations. Similarly total scores, i.e., CE plus discrimination scores, were multiplied by 8 and 7, respectively. Change or difference scores were obtained by subtracting Expression I scores from Expression II scores.

The judges were instructed by the experimenter with illustrations from a pilot study to clearly establish a standard. Two judges scored each paper independently and then compared results. The overall correlation between the two scores given each paper was .96. In the few cases where the judges did not agree, the experimenter acted as the final judge.

## RESULTS

Total difference scores did not differ significantly from group to group because CE scores and discrimination scores did not follow the same pattern. Consequently, total scores merely concealed information and can be ignored. Since the discrimination and CE scores in contrast to total scores appeared to differ markedly from group to group on Expression I, it seemed desirable to analyze Expression I scores and Expression II scores as well as change scores.

### *Expression I Scores*

The discrimination scores for the first pair of plants were substantially better for those that drew (see Tables 2 and 3). As would be expected, differences on Expression I according to Expression II code were not significant.

The number of children with CE scores of 1 or 2 is shown in Table 4. The proportion indicating cause was significantly greater among those who wrote on Expression I ($\chi^2 = 14.9$, $df = 1$, $p < .001$). For Expression I, it was easier to write about the cause of the differences but easier to indicate the differences themselves in a drawing.

TABLE 2

MEAN DISCRIMINATION SCORES BY EXPERIMENTAL GROUP

| Expression I code | Expression II code | | Total × Expression I code |
|---|---|---|---|
| | Write | Draw | |
| | On Expression I | | |
| Write | 19.1 | 16.4 | 17.7 |
| Draw | 21.8 | 23.2 | 22.5 |
| Total × Expression II code | 20.4 | 19.8 | |
| | On Expression II | | |
| Write | 16.4 | 10.9 | 13.6 |
| Draw | 14.8 | 14.8 | 14.8 |
| Total × Expression II code | 15.6 | 12.8 | |
| | Change score | | |
| Write | −2.7 | −5.5 | −4.1 |
| Draw | −7.0 | −8.4 | −7.7 |
| Total × Expression II code | −4.8 | −7.0 | |

TABLE 3

ANALYSES OF VARIANCE OF DISCRIMINATION SCORES

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Expression I | | | |
| Expression I code (A) | 1 | 501.1 | 12.25** |
| Expression II code (B) | 1 | 10.2 | .25 |
| A × B | 1 | 92.0 | 2.25 |
| Within groups | 84 | 40.9 | |
| Expression II | | | |
| Expression I code (C) | 1 | 28.4 | .06 |
| Expression II code (D) | 1 | 163.6 | 3.36 |
| C × D | 1 | 163.6 | 3.36 |
| Within groups | 84 | 48.7 | |
| Change scores | | | |
| Expression I code (E) | 1 | 290.9 | 4.48* |
| Expression II code (F) | 1 | 92.1 | 1.42 |
| E × F | 1 | 10.2 | .16 |
| Within groups | 84 | 64.9 | |

* $p < .05$.
** $p < .01$.

TABLE 4

NUMBER OF SUBJECTS EXPRESSING CE BY EXPERIMENTAL GROUP

| Expression I | Expression II code | | Total × Expression I code |
|---|---|---|---|
| | Write | Draw | |
| | On Expression I | | |
| Write | 16 | 13 | 29 |
| Draw | 4 | 7 | 11 |
| Total × Expression II code | 20 | 20 | |
| | On Expression II | | |
| Write | 12 | 5 | 17 |
| Draw | 9 | 7 | 16 |
| Total × Expression II code | 21 | 12 | |

### *Expression II Scores*

Mean discrimination scores for Expression II did not differ significantly among the experimental groups (Tables 2 and 3). On Expression II the number of subjects with CE scores of 1 or 2 (Table 4) was significantly greater among those who wrote than among those who drew ($\chi^2=3.9$, $df=1$, $p<.05$). Again it was easier to write about the cause than to show it in a drawing. Success in the expression of the cause of the differences on Expression II was independent of the code used on Expression I ($\chi^2=.1$, $df=1$).

### *Change Scores*

Most subjects did less well on Expression II in spite of the high degree of similarity between the two pairs of plants. Mean discrimination change scores (Expression II minus Expression I) were all negative (see Table 2) as were most of the mean CE change scores (not shown).

Among the discrimination change scores only the effects of Expression I were significant with the two groups who drew first showing greater loss than those that wrote first (see Table 3). If shifting code as such created difficulties one would expect a significant interaction. Thus, the hypothesis was not supported for the discrimination scores.

The variance analysis for CE change scores (not shown) indicated that only Expression I main effects were significant with the draw first groups showing a gain and the write first groups showing a loss. However, since CE scores turned out to be largely an all-or-none matter, the analysis of variance of these scores is misleading. The data shown in Table 5 indicate why this is so. Fifty-nine of the 88 subjects received the same CE score both times. Almost all of those who did indicate cause could not improve since their scores were already 2. Most of the remainder had 0 scores so loss was not possible. In Table 5 the

TABLE 5

RELATIVE FREQUENCY OF CE SCORE GAINS AND LOSSES

| Group | Number who could | | Number who did | | % number possible | | % difference | $\chi^2$ |
|---|---|---|---|---|---|---|---|---|
| | Gain | Lose | Gain | Lose | Gaining | Losing | | |
| WW | 7 | 16 | 3 | 7 | 43 | 44 | −1 | .2 |
| WD | 9 | 13 | 1 | 9 | 11 | 69 | −58 | 5.1* |
| DW | 18 | 4 | 6 | 3 | 33 | 75 | −42 | .9 |
| DD | 15 | 7 | 0 | 1 | 0 | 14 | −14 | .2 |

* $p<.03$.

percentage of subjects who did gain of those that could gain is compared with the percentage of subjects that did lose in score of those who could so lose. It can be seen that in the WW group, for example, 3 gained and 7 lost in score but the percentages are about the same.

A chi square comparing these two percentages was made for each group. The hypothesis under test in each case refers to whether the percentage gaining differs from percentage losing. Of the four groups only the WD group had significantly more losers than gainers. Perhaps because of this group the combined WD and DD groups showed a significantly greater proportion losing than gaining ($\chi^2=9.9$, $df=1$, $p<.005$) as did the combined DW and WD groups ($\chi^2=6.8$, $df=1$, $p<.01$).

It may be noted that a replication of this experiment in a single third grade class whose mean IQ was about 130 yielded parallel results.

## DISCUSSION

On Expression I, it was easier to draw the differences and easier to write about the cause of these differences. It is not always easy to describe in words variations in, say, the amount of wilt shown by a leaf, but it is not difficult to draw a droopy leaf. Similarly there is no convenient drawing symbol to express the word "because." Many subjects reported that they knew why the plants were not alike but could not express this in their drawings.

The substantial loss in score on Expression II is in some ways curious. One might have expected that the task would have been easier the second time since the plants were so similar. Possibly repeating the task was not challenging. However losses in discrimination scores were just as great among those changing form as those repeating the same task. Since most subjects had a rest period between expressions, fatigue should not have affected the scores very much.

The most reasonable explanation for the large losses in discrimination scores may be inferred from the prominence of the bloom on the begonia plant. Very few subjects failed to note this difference and it may have overshadowed the other differences, which having been noted once, no longer seemed important.

This coupled with the fact that higher discrimination scores were obtained by the group that drew first may explain why the only treatment effect found among the discrimination change scores was the greater loss in these groups. This would mean that the discrimination score phenomena were largely a reflection of the nature of the particular materials used.

However, CE scores did not follow this pattern. To be sure, the relative ease of writing about causation probably explains the more frequent success among the writing groups. But CE change scores do not permit any explanation based on code used in Expression I. From the data on CE change scores one can infer either (*a*) that shifting code, as hypothesized, inhibits expression, or (*b*) that for some unknown reason drawing on second expression (but not first) interfered with CE expression, or (*c*) that writing inhibits drawing but drawing does not inhibit writing. Strictly speaking the data do not require rejection of any of these three possibilities, but since three of the four subjects in the draw-write group who could lose did so, the third explanation seems improbable.

While the significant loss in CE score for the two groups shifting code (Groups WD and DW combined) may be largely because of the WD group, the significantly more frequent loss found among the subjects who drew on Expression II (Groups WD and DD combined) is almost entirely because of the

WD group alone. Thus the second explanation also seems improbable.

Although a check of the hypothesis is needed in a situation where an adequate range of change scores is possible, the present data are more consistent with the notion that shifting code inhibits the expression of inductively derived generalizations than they are with the two other possibilities.

It should be noted again that the discrimination scores showed no such effect so any inference that shifting code inhibits expression generally is not justified. Perhaps the amount of interference is related to the complexity of the concept. The simpler concepts would show little or no interference while the more complex concepts would show maximum interference.

Thus without denying the possible adequacy of other explanations we feel that the hypothesis about the effect of shifting code of expression is a worthy competitor of the proposition that something peculiar to the process of verbalizing principles interferes with the ability to apply these principles. In other words the difficulties encountered by the subjects in the investigations of Hendrix and Schwartz in particular may have been a consequence of the shift in code rather than the effect of verbalizing as such.

Furthermore giving a rule in words to subjects who then must work with essentially nonverbal materials (e.g., Katona) may simply complicate the task by adding the feature of code translation, especially when skill in translation has not yet developed.

It seems possible that this notion might reduce some of the apparent contradictions in the literature on instruction and discovery in the learning and transfer of generalizations and concepts.

## REFERENCES

BUSWELL, G. T., & KERSH, B. Y. Patterns of thinking in solving problems. *U. Calif. Publ. Educ.,* 1956, **12**, 63–148.

CORMAN, B. R. The effects of varying the amounts and kinds of information as guidance on problem solving. *Psychol. Monogr.,* 1957, **71**(2, Whole No. 431).

HENDRIX, GERTRUDE. A new clue to transfer of training. *Elem. Sch. J.,* 1948, **48**, 197–208.

KATONA, G. *Organizing and memorizing.* New York: Columbia Univer. Press, 1940.

KERSH, B. Y. The adequacy of "meaning" as an explanation for the superiority of learning by independent discovery. *J. educ. Psychol.,* 1958, **49**, 282–292.

SCHWARTZ, B. D. Certain relations between verbalization and concept formation. The destruction of ideas by words: *Dissert. Abstr.,* 1955, **15**, 453. (Abstract)

SMOKE, K. L. An objective study of concept formation. *Psychol. Monogr.,* 1932, **42**(4, Whole No. 191).

SZEKELY, L. Knowledge and thinking. *Acta psychol., Amsterdam,* 1950, **7**, 388–407.

(Received January 18, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 4, 208–212

# NOTE ON THE APPLICATION OF FACTOR ANALYSIS TO THE STUDY OF FACULTY MORALE[1]

RICHARD C. RICHARDSON, JR., AND CLYDE E. BLOCKER
*University of Texas*

The purpose of this study was to apply the technique of factor analysis to the evaluation of faculty morale in a college setting. A Differential Morale Attitude Inventory was constructed and administered to 66 members of the faculty of a Midwestern junior college. The inventory was divided into 12 differential categories which discriminated among responses beyond the .01 level of significance when subjected to a principal axis factor analysis. After varimax rotation, 4 factors were tentatively identified: supervision, self-integration, institutional environment, and employment rewards. The 4 factors were then described and related to those appearing in the industrial setting.

It appears evident from an examination of the literature that the attention of educational administrators is being drawn increasingly to the area of staff morale. While it has not yet been convincingly demonstrated that there is any significant relationship between morale and productivity, there is general agreement that, quite apart from any effect morale might have on teaching effectiveness, high morale is desirable.

The most common approach to the evaluation of staff morale has been through the use of some sort of morale attitude inventory. Typically, through intuition, examination of the literature, or interviews, a series of statements are constructed and combined in an inventory. The staff members are then asked to select a response for each item with a favorable response being interpreted as a high level of morale and an unfavorable response as a relatively lower level.

Existing morale attitude inventories, patterned along this general line, exhibit a number of weaknesses. The majority of them yield only a general measure of morale. A differential survey which would provide an indication of specific conditions contributing to low morale would be much more useful to the administrator interested in taking remedial action.

Those inventories which attempt to be diagnostic fall into two general classes. One class, well represented by the work of Redefer (1959), treats each individual item as a factor and attempts to generalize on this basis. Such an approach overlooks the well-recognized lack of reliability of individual test items. A second approach groups items into categories on a more or less a priori basis, which results in a blurring of potential differences. Thus, the results reported by Chandler (1959) and Mathis (1959) indicated that individuals within the schools studied did not project differential feelings of morale into specific areas of their inventory. This finding, which is in contradiction to most industrial studies, can be best explained by an examination of the categories they used and their procedures for assigning items.

In studies conducted by industrial psychologists, the number of morale factors found has varied, depending upon the method of factor analysis used and whether the items were analyzed individually or as parts of subtests. Baehr and Renck (1950), using a combination of item and category analysis in three studies, reported five primary factors. These were organization and management, immediate supervision, material rewards, fellow employees, and job satisfaction.

---

[1] This paper is based upon a doctoral dissertation directed by the junior author and submitted by the senior author to the Graduate Council of the University of Texas, 1963.

## METHOD

A review of educational and industrial morale research revealed the existence of a number of differential areas into which most of the attitudes commonly associated with job satisfaction appeared to fit. On the basis of this review, 12 categories were hypothesized and defined:

1. Communication—The expressed attitude of staff members toward the degree of freedom with which information travels up and down the administrative structure.
2. Confidence in Administration—The expressed attitudes of staff members toward the fairness, consistency, competence, and effectiveness of administration beyond the immediate supervisory level.
3. Relations with Immediate Supervisor—The expressed attitudes of staff members toward the fairness, consistency, competence, and effectiveness of immediate supervisors.
4. Relations with Fellow Employees—The expressed attitude of staff members toward the friendliness, cooperativeness, and competency of persons holding positions of a similar level in the institution.
5. Relations with Students—The expressed attitude of staff members toward students and their activities.
6. Status and Recognition—The expressed attitude of staff members toward the respect accorded their position and the recognition they receive.
7. Identification with Institution—The expressed attitude of staff members toward their pride and interest in the institution and the degree to which they feel they belong.
8. Professional Growth and Advancement—The expressed attitudes of staff members toward their opportunities for professional growth and development and their feelings concerning the advancement possibilities of their position.
9. Adequacy of Salary—The expressed attitudes of staff members toward the adequacy of salary.
10. Adequacy of Fringe Benefits—The expressed attitudes of staff members toward the employee benefits offered by the institution.
11. Work Environment—The expressed attitudes of staff members toward working conditions in the institution.
12. Work Load—The expressed attitudes of staff members toward the work demands imposed by their position.

Several existing morale attitude inventories were examined and from among their combined contents 85 statements were selected on the basis of specific content and apparent relationship to the categories hypothesized. A group of 10 judges was asked to assign the statements to the hypothesized categories. Eleven statements were eliminated because of lack of agreement. The remaining 74 were combined in an attitude inventory and administered to 66 faculty members of a midwestern junior college.

The responses were scored and analyzed to determine if the faculty members revealed differential attitudes toward the 12 categories of the inventory. Using a modification of chi square reported by McNemar (1955, p. 232) it was discovered, contrary to the findings of Mathis (1959) and Chandler (1959), that the differential categories discriminated among the responses well beyond the .01 level of significance.

The 12 categories of the morale attitude inventory were subsequently subjected to a principal axis factor analysis utilizing the CDC 1604 computer at the University of Texas. The program, written by Earl Jennings and Donald J. Veldman, also yielded varimax rotation of factors.

## RESULTS

The intercorrelation matrix is presented in Table 1. The unrotated orthogonal factor loadings are given by Table 2. Table 3 indicates the factor loadings after normal varimax rotation.

It should be noted that varimax rotation as described by Kaiser (1958) presents an analytic solution which has been shown in a number of instances to approach the best orthogonal simple solution closely. One of the principal arguments advanced in favor of the varimax solution is that it removes the element of subjectivity from factor analysis and, thus, brings it more closely in line with the objective requirements of scientific inquiry.

TABLE 1

TABLE OF INTERCORRELATIONS

| Category | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | .789 | .437 | .342 | .225 | .501 | .343 | .432 | .171 | .296 | .383 | .307 |
| 2 | | | .503 | .338 | .359 | .522 | .341 | .512 | .060 | .237 | .477 | .397 |
| 3 | | | | .433 | .109 | .436 | .256 | .263 | .058 | .128 | .306 | .100 |
| 4 | | | | | .131 | .435 | .456 | .203 | .197 | .115 | .305 | .226 |
| 5 | | | | | | .389 | .395 | .298 | .005 | .245 | .371 | .350 |
| 6 | | | | | | | .616 | .377 | .163 | .204 | .498 | .333 |
| 7 | | | | | | | | .410 | .154 | .118 | .307 | .228 |
| 8 | | | | | | | | | .118 | .239 | .428 | .266 |
| 9 | | | | | | | | | | .397 | .241 | .198 |
| 10 | | | | | | | | | | | .365 | .177 |
| 11 | | | | | | | | | | | | .506 |
| 12 | | | | | | | | | | | | |

## DISCUSSION

It was possible to identify the factors revealed by this study and to relate them to the factors identified in the industrial studies of morale research reviewed in this paper.

Factor I has been tentatively identified as Supervision. The categories of the inventory that have high loadings on this factor are Communication, Confidence in Administration, Relations with Immediate Supervisor, and Professional Growth and Advancement. This factor appears to be associated with the attitudes of faculty members toward the overall effectiveness of administrative processes. This factor is comparable to two of the factors identified by Baehr and Renck (1950), Organization and Management and Immediate Supervision. The finding that a single factor in the educational environment appears to account for two factors in the industrial situation may be partially explained

TABLE 2

ORTHOGONAL FACTOR LOADINGS EXTRACTED BY THE PRINCIPAL AXIS METHOD[a]

| Category | Factor I | Factor II | Factor III | Factor IV | $h^2$ |
|---|---|---|---|---|---|
| 1. Communication | .75 | −.38 | .17 | −.14 | .75 |
| 2. Confidence in Administration | .80 | −.39 | −.04 | −.20 | .83 |
| 3. Relations with Immediate Supervisor | .56 | −.18 | .31 | −.50 | .69 |
| 4. Relations with Fellow Employees | .57 | .41 | .41 | −.22 | .71 |
| 5. Relations with Students | .53 | .16 | −.61 | .13 | .70 |
| 6. Status and Recognition | .78 | .27 | −.01 | −.14 | .69 |
| 7. Identification with Institution | .64 | .57 | −.06 | −.14 | .76 |
| 8. Professional Growth and Advancement | .64 | −.16 | −.16 | .01 | .46 |
| 9. Adequacy of Salary | .28 | .18 | .49 | .68 | .82 |
| 10. Adequacy of Fringe Benefits | .43 | −.23 | .22 | .59 | .63 |
| 11. Work Environment | .71 | −.07 | −.12 | .26 | .60 |
| 12. Work Load | .55 | −.01 | −.32 | .30 | .50 |

[a] Unity used as an estimate of communality.

TABLE 3

FACTOR LOADINGS AFTER VARIMAX ROTATION

| Category | Factor I | II | III | IV | $h^2$ |
|---|---|---|---|---|---|
| 1. Communication | .80 | .16 | −.20 | .19 | .75 |
| 2. Confidence in Administration | .82 | .13 | −.37 | .06 | .83 |
| 3. Relations with Immediate Supervisor | .73 | .37 | .10 | −.09 | .69 |
| 4. Relations with Fellow Employees | .28 | .78 | .05 | .16 | .71 |
| 5. Relations with Students | .04 | .16 | −.82 | −.07 | .70 |
| 6. Status and Recognition | .37 | .62 | −.41 | .07 | .69 |
| 7. Identification with Institution | .08 | .77 | −.40 | .00 | .76 |
| 8. Professional Growth and Advancement | .46 | .13 | −.46 | .11 | .46 |
| 9. Adequacy of Salary | −.10 | .22 | .01 | .87 | .82 |
| 10. Adequacy of Fringe Benefits | .23 | −.08 | −.23 | .72 | .63 |
| 11. Work Environment | .36 | .18 | −.56 | .35 | .60 |
| 12. Work Load | .16 | .07 | −.64 | .23 | .50 |

by an examination of obvious differences between the organization of the two environments. The hierarchical structure of the industrial concern emphasizes the role of the immediate supervisor. In the college environment, the hierarchical structure is much more limited in keeping with the concept that professionals are capable of self-direction and require a minimum of coordination.

Factor II has been termed Self-Integration. Relations with Fellow Workers, Status and Recognition, and Identification with the Institution all have high loadings on this factor. The image that a faculty member has of himself in relation to his colleagues and the institution, as well as the degree of respect with which he views his colleagues, are characteristic of the categories with high loadings on Factor II. The most probable explanation for the broad character of Self-Integration as contrasted with Baehr and Renck's (1950) related factor, Fellow Employees, appears to be the fact that, within a colleague group of professionals, more support is derived from peers than from the administration. Consequently, a professional's self-image and the degree to which he identifies with his institution are more closely related to his views of his colleagues and their views of him than would be likely in the industrial environment among nonprofessionals.

Factor III has been called Institutional Environment. Categories with high loadings on this factor are Relations with Students, Professional Growth and Advancement, Work Environment, and Work Load. Factor III is best thought of as relating to the conditions under which a faculty member works. Several of the categories which have high loadings on this factor also have substantial loadings on Factor I. It is likely that many of the conditions loading on this Factor are also related to Supervision, but their relationship is of an indirect nature. For example, the work load of a given faculty member is determined by administrative procedures, but policy ordinarily governs this assignment and it is probable that a given level is accepted as a part of the conditions of employment, rather than being considered as within the scope of administrative judgment. Factor III closely parallels Baehr and Renck's (1950) factor, Job Satisfaction.

Factor IV has been named Employment Rewards. Two categories, Adequacy of Salary and Adequacy of Fringe Benefits, have high loadings on this factor. The general nature of this factor is obvious from the categories which load on it. Factor IV is directly comparable to Baehr and Renck's (1950) factor, Material Rewards. It is interesting to note that the two categories defining Factor IV have small loadings on the other three factors. The majority of studies in industrial research have found that the area of job rewards was a relatively independent dimension of employee morale.

## CONCLUSIONS

The findings of this study indicate the desirability of utilizing the recent advances in industrial research to identify and describe the basic dimensions of morale in the educational environment.

While similarities between the findings of this study and those of industrial research were emphaized, of equal importance is the clarification of basic differences that must result from dissimilarities in the organization and purpose of educational institutions as opposed to business concerns. The technique of factor analysis can provide the opportunity to improve both the methods used in assessing faculty morale and the clarity with which these assessments are reported to educational administrations.

## REFERENCES

BAEHR, MELANY E., & RENCK, R. The definition and measurement of employee morale. *Adm. sci. Quart.*, 1950, **3**, 157–184.

CHANDLER, B. J. Salary policies and teacher morale. *Educ. Admin. Superv.*, 1959, **45**, 107–110.

KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, **23**, 187–200.

MATHIS, C. The relationship between salary policies and teacher morale. *J. educ. Psychol.*, 1959, **50**, 275–279.

MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1955.

REDEFER, F. L. Factors that affect teacher morale. *Nation's Sch.*, 1959, **63**, 59–62.

RICHARDSON, R. C., JR. Staff morale: An investigation involving the development and testing of a Differential Morale Attitude Inventory. Unpublished doctoral dissertation, University of Texas, 1963.

(Received January 13, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 4, 213–216

# CONCEPT ATTAINMENT OF HIGH SCHOOL SOPHOMORES[1]

LeROY A. OLSON[2]
*University of Wisconsin*

Concept attainment problems involving pictorial and symbolic stimuli were presented to 96 high school sophomore students. Type of instance, order of instance presentation, sex, and amount of practice were included in a complete factorial design. Analyses of variance indicated: attainment of pictorial instance concepts was significantly higher when stimuli were presented in an obvious to subtle order; no significant differences in concept attainment when positive instances were presented alone or in conjunction with negative instances; Ss with less practice did as well, after the first problem, as more experienced Ss; and sex differences in concept attainment problems appeared to depend on the nature of the concept.

One type of conceptualizing activity, known as concept attainment, involves distinguishing examples of a class of things which one wishes to discriminate (Bruner, Goodnow, & Austin, 1956). This process appears to play an important part in the classroom activity of the secondary school. An understanding of the effects on the concept attainment process of variables which may be manipulated in the classroom might help instructors to facilitate such a process.

A number of variables which way be controlled in classroom presentations were investigated. The specific purpose of this study was to determine the effects of varying the type of stimuli and the order of presentation of stimuli on the concept attainment of high school sophomores.

Two types of stimuli were used: positive instances and negative instances. Positive instances were examples of the concept to be attained while negative instances were examples of things which did not illustrate the concept. The first hypothesis to be tested was that subjects attain concepts with equal efficiency when only positive instances are presented as when some positive instances and some negative instances are presented.

A second hypothesis to be tested was that subjects attain concepts equally well when the stimuli or instances are presented in an obvious to subtle order, in a random order, or in a subtle to obvious order.

A third hypothesis that subjects who had more practice in attaining concepts would attain concepts with the same efficiency as subjects with a small amount of practice was also tested.

Finally, the hypothesis that boys and girls were equally efficient in attaining concepts was tested.

## METHOD

### *Selection of Subjects*

The subjects for the study were selected at random from the sophomore class of a public secondary school. They ranged in age from 15 to 17 years, with an average age of 15.5 years. The average Henmon-Nelson group intelligence test raw score was 45.0. A subject 15.5 years old who received a raw score of 45 would have an IQ of 98, based on Henmon-Nelson national norms.

---

[1] The data on which this paper is based were included in the author's dissertation, presented in partial fulfillment of the requirements for the PhD at the University of Wisconsin. The author is indebted to C. W. Harris, major advisor, for his advice and encouragement.

[2] Now at Michigan State University.

### *Statistical Design*

The study utilized a complete factorial arrangement of experimental groups (Lindquist, 1953). Subjects were presented with four problems, each of which was analyzed separately. The complete factorial design permitted the use of the normal theory analysis of variance with fixed effects. Four subjects were randomly assigned to each treatment group in order to control extraneous variables and to increase the precision of the experiment.

The first two problems were presented to 48 subjects in a 2×2×3 factorial design employing two levels of type of instance (only positive instances, and both positive and negative instances), boys and girls, and three levels of order of instance presentation (obvious to subtle order, random order, and subtle to obvious order).

The remaining two problems were presented to 96 subjects. The resulting 2×2×2×3 factorial design involved two levels of practice (those who had completed the first two problems, and those who had not), two levels of type of instance (only positive instances, and both positive and negative instances), boys and girls, and three levels of order of instance presentation (obvious to subtle order, random order, and subtle to obvious order).

### *Stimuli*

Four concept attainment problems were employed: (*a*) objects associated with reciprocating motion, (*b*) levers, (*c*) groups of five letters with three letters in alphabetical order, and (*d*) numbers which were multiples of three. An independent sample of secondary public school students was asked to classify 50 examples of the first two concepts as obvious, medium, or subtle examples of these concepts. With the help of these ratings, 10 stimuli were selected as instances of each concept. Of the 10, 3 had been judged as obvious examples of the concepts under consideration, 4 had been judged to be of medium subtlety, and 3 were judged to be subtle examples of the concept. The stimuli for the alphabetical order problem and the multiple of three problem were selected by the experimenter. Five negative-instance stimuli were selected for each of the four concepts by the experimenter.

### *Presentation of Stimuli*

The pictures, letters, and numbers illustrating the stimuli for the four concepts were drawn in black on 11×15 inch white cards. The stimuli were presented at 30-second intervals. When a stimulus had been presented, it remained in the subject's view until all stimuli had been presented for a given problem. This procedure was intended to minimize the effects of visual discrimination and memory ability. Stimuli were presented to groups of subjects having identical experimental characteristics in order to maintain a classroom atmosphere.

### *Evaluation of Responses*

After seeing each new stimulus, the subjects were asked to indicate on an answer sheet what they considered the common element or concept to be. A subject was considered to have attained a concept when he indicated the correct common element and continued to indicate the same common element after seeing each remaining stimulus. A subject's score on a problem was inversely related to the number of stimuli viewed before attaining the concept. Thus a subject attaining a concept after seeing two stimuli received a score of 10 and a subject viewing all 10 stimuli before discovering the common element was assigned a score of 1.

## RESULTS

The data for each of the four concept attainment problems were analyzed separately. The statistical significance of each of the experimental factors was determined by the analysis of variance. For the reciprocating motion concept attainment problem, the order of instance presentation factor was statistically significant at beyond the .001 level ($F = 13.28$, $df = 2/36$). No other mean squares were statistically significant. Means and standard deviations for main effect groups are shown in Table 1.

The order of instance presentation factor was also statistically significant (at the .05 level) for the lever concept attainment problem ($F = 3.28$, $df = 2/36$). In addition, the sex factor was statistically significant at the .03 level for this problem ($F = 5.74$, $df = 1/36$). No other mean squares were statistically significant.

The analysis of scores on the multiple-of-three concept attainment problem re-

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR MAIN EFFECT GROUPS BY PROBLEM

| Main effect | Reciprocating motion | | Lever | | Multiple of three | | Alphabetic order | |
|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Sex | | | | | | | | |
| Male | 4.5 | 3.7 | 4.6 | 4.4 | 6.6 | 4.0 | 4.4 | 4.4 |
| Female | 4.4 | 4.1 | 2.2 | 2.4 | 7.5 | 4.1 | 3.9 | 4.1 |
| Type of instance | | | | | | | | |
| Positive only | 4.9 | 4.1 | 3.1 | 3.6 | 7.6 | 4.2 | 4.9 | 4.4 |
| Positive and negative | 4.0 | 3.7 | 3.7 | 3.8 | 6.5 | 3.9 | 3.4 | 3.9 |
| Order of presentation | | | | | | | | |
| Obvious to subtle | 7.8 | 3.6 | 5.1 | 4.6 | 7.1 | 4.1 | 4.8 | 4.5 |
| Random | 3.3 | 3.2 | 3.1 | 3.1 | 7.8 | 4.2 | 4.6 | 4.3 |
| Subtle to obvious | 2.3 | 2.3 | 2.0 | 2.6 | 6.2 | 3.9 | 3.1 | 3.7 |
| Amount of practice | | | | | | | | |
| Practice | | | | | 7.9 | 3.7 | 4.2 | 4.3 |
| No practice | | | | | 6.1 | 4.2 | 4.1 | 4.2 |

vealed a statistically significant difference (at the .03 level) in favor of subjects who had completed two problems ($F=5.38$, $df=1/72$). In addition, the second order interaction between sex, type of instance, and order of instance presentation was statistically significant at the .04 level ($F=3.45$, $df=2/72$). No other mean squares were statistically significant.

The analysis of scores on the alphabetic order concept attainment problem indicated that none of the factors under consideration produced statistically significant results.

## DISCUSSION AND CONCLUSIONS

The hypothesis that subjects attain concepts with equal efficiency when only positive instances are presented as when both positive and negative instances are presented could not be rejected for any of the four concept attainment problems. Subjects receiving only positive instances obtained higher scores for three of the four problems but in no case were the differences between groups receiving different types of instances statistically significant. This is contrary to the results obtained by Hovland and Weiss (1953), although the materials and methods of presentation used in this study differ considerably from those in the earlier study. The results agree, however, with those obtained by Smoke (1933).

The hypothesis that subjects attain concepts equally well when instances are presented in an obvious to subtle order was rejected for the reciprocating motion and lever concept attainment problems. The subjects receiving instances in obvious to subtle order achieved significantly higher scores on these pictorial material problems. Underwood and Richardson (1956) achieved similar results with verbal materials. The above hypothesis could not be rejected for the symbolic material problems: multiples of three and letters in alphabetic order.

It was not possible to determine whether the method of instance scaling or the type of material used was responsible for the failure of an obvious to subtle order of stimulus presentation to result in more efficient attainment of symbolic concepts as it did for pictorial concepts.

The hypothesis that subjects with more practice in solving concept attainment problems would attain concepts with the same efficiency as subjects with a small amount of practice was rejected for the mutliple-of-three problem but could not be rejected for the alphabetic order problem. This result indicated that unpracticed subjects attained concepts less well on their first problem in comparison with more experienced subjects, but that they attained concepts as well on their second problem as more experienced subjects. Thus, although a practice effect existed, it was of relatively short duration.

The hypothesis that boys and girls attain concepts equally well could not be rejected for the reciprocating motion, multiple-of-three and alphabetic order problems. This hypothesis was rejected, however, for the lever concept attainment problem, on which boys scored significantly higher than girls. Whether boys or girls are more efficient at solving a concept attainment problem appeared to depend on the nature of the particular concept.

All subjects received the problems in the same order, permitting the use of a prepared test answer booklet and a group test procedure, as well as avoiding the confounding of main effects in the analysis. It was found that concepts can be effectively attained when the instances are presented to subjects in a group situation.

## REFERENCES

Bruner, J. S., Goodnow, Jacqueline J., & Austin, G. A. *A study of thinking.* New York: Wiley, 1956.

Hovland, C. I., & Weiss, W. Transmission of information concerning concepts through positive and negative instances. *J. exp. Psychol.,* 1953, **45,** 175–182.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

Smoke, K. L. Negative instances in concept learning, *J. exp. Psychol.,* 1933, **16,** 583–588.

Underwood, B. J., & Richardson, J. Some verbal materials for the study of concept formation. *Psychol. Bull.,* 1956, **53,** 84–95.

(Received January 13, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 4, 217–226

# FURTHER VALIDATION OF THE ENVIRONMENTAL ASSESSMENT TECHNIQUE[1]

ALEXANDER W. ASTIN
*National Merit Scholarship Corporation, Evanston, Illinois*

Environmental Assessment Technique (EAT) measures were computed for 76 colleges and universities from published data about size, intelligence level, and major fields of study of students at each institution, and validation data were obtained from seniors. 18 items of a questionnaire completed by the latter described their college environments and 21 items described the perceived "effects" of 4 college years. Mean responses to 14 of the 18 college environment items were significantly related to EAT variables in the predicted direction; 15 of 21 perceived effects of college were significantly related to at least one EAT variable. Student intelligence accounted for a larger proportion of variance in the validation items than any other EAT variable.

The Environmental Assessment Technique (EAT) is a method for assessing the college environment based on eight attributes of the student body: the Size of the Student Body, the Intelligence Level of the students, and the "personal orientations" of the students as indicated by the percentage of students in each of six classes of major fields (Realistic, Intellectual, Social, Conventional, Enterprising, and Artistic). With the possible exception of the Intelligence Level of the Student Body, data for computing EAT measures for any institution can usually be obtained from published sources, thus eliminating the necessity of relying on informants at the institution.

In an earlier report on a sample of 36 colleges (Astin & Holland, 1961), EAT variables were found to have substantial correlations with many of the scales of the College Characteristics Index (CCI; Pace & Stern, 1958). Also, the percentage of students majoring in each of the six groups of fields was found to be highly reliable over intervals of 1–6 years.

The primary purpose of the present study was to test and extend the construct validity of the EAT using a larger sample of institutions. The correlations found earlier between EAT measures and CCI scales—although relatively high—did not seem satisfactorily to establish the construct validity of the EAT, since many CCI items inquire directly into the student's major fields of study and related interests. For example, the CCI Pragmatism scale, which correlated .73 with the EAT Realistic Orientation (Astin & Holland, 1961), includes the following items: "Most students are interested in careers in business, engineering, management, and other practical affairs" (True); and "Students are more interested in specialization than in general liberal education" (True). Since the percentage of students majoring in engineering determines to a great extent the Realistic Orientation of the college, it is hardly surprising that these two scales are significantly related. A similar example occurs with the CCI scale, Sentience, which includes the following items: "There is a lot of interest here in poetry,

---

[1] This study is part of the research program of the National Merit Scholarship Corporation and was supported by a grant from the National Science Foundation.

music, painting, sculpture, architecture, etc." (True); and "The college has invested very little in drama and dance" (False). Sentience, in turn, correlated .69 with EAT Artistic Orientation, which is defined as the percentage of students majoring in art, music, English, journalism, etc.[2]

Since CCI is, in a sense, an indirect method for determining the distribution of major fields at an institution, then correlations between CCI scales and EAT personal orientations may be mediated by the curriculum at the institution, rather than by aspects of its "psychological climate." In short, high correlations between two methods for counting the major fields at an institution do not necessarily indicate that either method is a valid measure of the psychological environment at the college.

To avoid these difficulties, the present study attempted to explore the construct validity of EAT directly by using items specifically designed to test the proposed interpretations of each EAT variable.

A secondary purpose of the present study was to try out an alternative method for scoring the six EAT personal orientations. In the earlier report (Astin & Holland, 1961), each major field was treated as a "pure" type, i.e., as belonging to only one of the six personal orientations. However, since certain major fields probably combine attributes of more than one personal orientation, it seemed likely that more valid estimates of each personal orientation could be secured by applying appropriate differential weights to these major fields. Accordingly, in the present study the pure (unit weight) method was compared with a method of differential weighting to see which yielded the more accurate scores for the six EAT personal orientations.

## METHOD

### *Sample of Institutions*

Information about the college environments was obtained from students at some 76 different colleges and universities which enrolled 15 or more students from a larger sample of Finalists and recepients of the Letter of Commendation in the 1957 National Merit Scholarship program. The students, who numbered from 15 to 216 per college, completed a mailed questionnaire at approximately the end of their senior year in college.[3]

The EAT measures for each institution were computed in the manner outlined in the earlier report (Astin & Holland, 1961). In addition, the six EAT personal orientations were computed according to a system of differential weights which assigned certain major fields to two personal orientations. A given major field could thus be divided between two personal orientations using one of the following pairs of weights: .25 and .75, .50 and .50, and .75 and .25. (Major fields which were still considered pure types—i.e., belonging in a single personal orientation—were given a weight of 1.00 for that orientation.) Decisions about how to weight each major field were made primarily on the basis of Holland's (1962) theoretical formulations for each of the six personal orientations. Some of the more important changes from the pure system are as follows: students majoring in architecture (formerly scored only under the Intellectual Orientation) were weighted .50 for the Intellectual and .50 for the Artistic Orientations; students majoring in business or economics (formerly scored only as Conventional) were weighted .50 for the Conventional and .50 for the Enterprising Orientations; students majoring in sociology or psychology (formerly scored only as Social) were weighted .25 for the Intellectual and .75 for the Social Orientations.[4] Since the total weight contributed by any one student always summed to 1.00, the system of differential weights still resulted in six scores which summed to 100 (%) for each institution.

---

[2] Similar possible circularities occur with other EAT-CCI correlations, even those involving the Intelligence Level of the Student Body. For example, the Achievement scale of the CCI, which correlated .64 with Intelligence, includes the following items: "It is fairly easy to pass most courses without working very hard" (False); "the competition for grades is intense" (True). Thus, the Achievement scale may be an indirect method for determining the level of intelligence of the student body.

[3] This sample has been described in more detail elsewhere (Astin, 1963).

It was anticipated that the criterion used to select institutions for this study (i.e., that the college enroll at least 15 of these highly able students) would produce a sample in which the correlation between Size and Intelligence Level of the Student Body would be spuriously high, since any small institution, in order to be included, would have to enroll a student body of very high ability level. This expectation was borne out: in the sample of 76 colleges there was a correlation of —.57 between Size and Intelligence, as compared with a correlation of only —.09 found earlier for a less select sample of 335 institutions (Astin & Holland, 1961).

## *Questionnaire Items*

Two kinds of items were included in the questionnaire. The first group, which consisted of 18 items, was developed to examine the construct validity of each EAT variable. In writing these items, we tried to avoid merely restating the operations involved in scoring the variable. For example, it would probably be a relatively easy matter to "validate" the Conventional Orientation using items such as "Many of the students are planning careers in business" (True or False?), simply because the score for the Conventional Orientation is based largely on the percentage of students who are majoring in business. Similarly, it would be easy to demonstrate the high validity of the EAT variable, Size of the Student Body, by using an item such as "This is a very large campus."

To avoid such circularity, an attempt was made to design items which had only remote connections with the actual operations involved in scoring the EAT variable, but which were at the same time relevant to the proposed interpretations of the variable. For instance, to test the notion that the Conventional Orientation fosters conformity (Astin & Holland, 1961, p. 310), the following item was used: "There is a set way to dress which most students go along with."

The subject answered each item in this first group according to a 4-point scale: true (4), somewhat true (3), somewhat false (2), and false (1). The 18 items, together with summaries of the environmental constructs proposed for each EAT variable, are given below with their respective predicted relationships to tendency to respond "true" (negative or positive).

*Size of the Student Body.* Large institutions will be characterized by an impersonal atmosphere, with personal contacts between faculty and students at a minimum. The large number of students will probably result in a variety of highly organized student subcultures and a fairly well-defined status hierarchy of social groups.

Item 1. There is a lot of personal contact between students and faculty members (negative).

Item 2. Many of the social groups on campus have a definite "snob" appeal (positive).

*Ability Level of the Student Body.* Colleges with student bodies of relatively high intelligence will also be characterized by a relatively high emphasis on academic pursuits and on scholarship. Since ability level tends to be related to other assets of the institution, including the level of training of the faculty (Astin, 1962), we might also expect the quality of instruction at the institution to be related to the ability level of the student body.

Item 3. The typical student spends a lot of time in the library (positive).

Item 4. Many of the students are more interested in getting a degree than they are in what they will learn in college (negative).

Item 5. The quality of teaching and instruction is markedly superior to that in the average institution (positive).

*Realistic Orientation.*[5] The realistic environment will be characterized by a relatively high incidence of aggressive, masculine activities. "Acting out" (as opposed to "thinking through") will be the preferred method for coping with problems.

Item 6. Practical joking among the students is fairly common (positive).

---

[4] A complete table of the differential weights for scoring the EAT personal orientations has been deposited with the American Documentation Institute. Order Document No. 7622 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Remit in advance $1.25 for microfilm or $1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

[5] Theoretical formulations for the six EAT personal orientations were abstracted from the earlier paper (Astin & Holland, 1961) and from Holland's (1962) theory of vocational choice.

Item 7. An outburst or demonstration among the students would not be very likely (negative).

*Intellectual Orientation.* The intellectual environment will be characterized by a tendency to underemphasize interpersonal relations. One would expect the introverted student to feel more at home in this environment.

Item 8. A student who is introverted or stays pretty much to himself would feel out of place on this campus (negative).

Item 9. Many of the students are interested primarily in getting married and raising families (negative).

*Social Orientation.* The social environment will foster a relatively large amount of structural social activity. In particular, one would expect the acquisition of social skills to be encouraged.

Item 10. There is a good deal of emphasis on parties and dating (positive).

Item 11. A student who is not very skilled in etiquette or social graces would probably feel out of place on the campus (positive).

*Conventional Orientation.* The conventional environment will be characterized by a relatively high degree of conformity among the students and by an authoritarian attitude on the part of the faculty and administration.

Item 12. There is a set way to dress which most students go along with (positive).

Item 13. The administration exerts a lot of pressure on the students to conform to the rules and regulations (positive).

Item 14. Faculty members usually don't like the student to question their judgment or point of view (positive).

*Enterprising Orientation.* The enterprising environment will foster competition for power and support the development of verbal and persuasive skills.

Item 15. Students are always ready to argue or debate almost any issue (positive).

Item 16. There is plenty of competition among students for leadership positions on campus (positive).

*Artistic Orientation.* The artistic environment will be characterized by a highly individualized (or opposed to a group) approach to problem solving. One would expect the development of gross physical skills to be discouraged.

Item 17. Students are more inclined to pursue their own individual projects than to engage in group activities (positive).

Item 18. The students show a great deal of interest in competitive sports (negative).

The second group of items included in the questionnaire was intended to examine the construct validity of EAT on a more exploratory basis, rather than to test specific interpretations of the variables. Two of these 21 items were designed to elicit the student's general evaluation of his college: "What would you describe as the main *assets* or strong points of your particular college?" (open-ended response); and "What would you describe as the main *liabilities* or faults of your particular college?" (open-ended response).

The remaining 19 items were designed to assess the student's evaluation of how his 4 years in college had affected his values, interests, and personal development. The student was instructed to judge the extent to which the college had influenced him in the following areas: Clarity of vocational plans, study habits, poise in groups, general personal development (4-point scales: marked improvement, some improvement, no change, worse than when I entered college); interest in art and music, knowledge of etiquette, and social graces, ambition, interest in sports, concern over grooming and personal appearance, amount of drinking, interest in world affairs, concern for feelings of others, interest in science, self-confidence, tendency to argue and debate issues (4-point scales: marked increase, some increase, no change, less than when I entered college); attitudes toward formal religion, fraternities and sororities, marriage (5-point scales: much more positive, somewhat more positive, no change, somewhat more negative, much more negative); political attitudes (5-point scales: much more conservative, somewhat more conservative, no change, somewhat more liberal, much more liberal).

### *Analyses of Data*

Mean scores on each of the 39 items were computed separately for each of the 76 institutions. Product-moment correlations were then computed between these means and the 14 EAT variables (Size, Intelligence, and the six personal orientations scored in two ways).

Intercorrelations among the eight EAT variables scored by the method of differential weights were also computed using the original sample of 335 institutions (Astin & Holland, 1961).

## RESULTS

A comparison of the analyses indicated that of the two methods for scoring the six personal orientations, the method of differential weights is gen-

TABLE 1

INTERCORRELATIONS OF EAT VARIABLES USING TWO METHODS[a] OF SCORING THE SIX PERSONAL ORIENTATIONS
($N$ = 335 institutions)

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1. Size | | −.09 | .26 | −.28 | −.06 | .26 | −.25 | −.29 |
| 2. Intelligence | −.09 | | −.08 | .41 | −.36 | −.27 | .48 | .34 |
| 3. Realistic Orientation | .25 | −.06 | | −.24 | −.37 | −.17 | −.46 | −.57 |
| 4. Intellectual Orientation | −.16 | .47 | .24 | | −.40 | −.17 | .38 | .16 |
| 5. Social Orientation | −.17 | −.29 | −.53 | −.57 | | −.23 | −.32 | −.01 |
| 6. Conventional Orientation | .33 | −.33 | −.10 | −.25 | −.30 | | −.12 | −.30 |
| 7. Enterprising Orientation | −.09 | .32 | −.48 | .06 | −.29 | .29 | | .45 |
| 8. Artistic Orientation | −.26 | .13 | −.60 | −.30 | .51 | −.42 | −.03 | |

Note.—$r_{.05} = .11$, $r_{.01} = .14$.

[a] Intercorrelations using the differential weighting system for scoring the six personal orientations are given in the lower-left triangular matrix; the upper-right triangular matrix shows the intercorrelations based on the unit-weighting (pure) scoring system used in the previous report (Astin & Holland, 1961, Table 5).

erally superior to the method of unit weights used in the earlier study (Astin & Holland, 1961). Six of the 13 items chosen to test specific interpretations of the personal orientations yielded significant results in the predicted direction when the unit weights were used, whereas 9 of the 13 items yielded significant results when differential weights were used (see below).

Table 1 shows the intercorrelations—based on both scoring methods—of all eight EAT variables, for the original sample of 335 institutions (Astin & Holland, 1961). Using differential weights increases the amount of experimental dependence among the six personal orientations, but only slightly: the mean intercorrelation (ignoring sign) increases from .29 to .33. The intercorrelations which were most affected by the differential weighting system included: Realistic × Intellectual (from −.24 to .24), Intellectual × Enterprising (from .38 to .06), Intellectual × Artistic (from .16 to −.30), Social × Artistic (from −.01 to .51), Conventional × Enterprising (from −.12 to .29), and Enterprising × Artistic (from .45 to −.03). With the possible exception of the correlation between the Intellectual and the Artistic Orientations, the correlations produced by the differential weighting system conform more closely to the theoretical descriptions of the personal orientations (Holland, 1962). Correlations of the six personal orientations with Size and Intelligence were not changed appreciably by the differential weighting system, although the mean correlation (ignoring sign) was reduced slightly from .28 to .24.

Distributions of scores for the 335 institutions on each personal orientation also tend to be more symmetrical when the scores are based on differential weights than when they are based on unit weights.

Since using differential weights for the personal orientations resulted in scores which have greater validity and more desirable statistical properties than the scores based on unit weights, the remaining results are presented only for this revised scoring procedure.

Table 2 shows the correlations between the eight EAT variables and the mean scores for the 39 items.

Fourteen of the 18 predicted relationships (coefficients in boldface type in Table 2) were confirmed ($p < .01$). Of the 144 correlations computed between

TABLE 2

CORRELATIONS[a] OF EAT VARIABLES WITH 18 RATINGS OF THE COLLEGE ENVIRONMENT AND WITH 21 PERCEIVED EFFECTS OF THE COLLEGE
($N = 76$ Colleges, 15–216 student raters per college)

| Item | EAT variable | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Size of student body | Intelligence level of student body | Personal orientation | | | | | |
| | | | Realistic | Intellectual | Social | Conventional | Enterprising | Artistic |
| Ratings of the environment | | | | | | | | |
| 1. Amount of faculty-student contact | −74 | 35 | −53 | | 35 | −40 | 41 | 47 |
| 2. "Snob" appeal of social groups | 57 | −43 | | −33 | | 44 | | |
| 3. Time spent by students in library | −57 | 63 | −52 | | 37 | −50 | 50 | 59 |
| 4. Students' orientation to getting a degree rather than learning | 63 | −73 | 37 | −31 | | 56 | −30 | −44 |
| 5. Superior quality of teaching | −55 | 74 | −35 | | | −59 | 42 | 41 |
| 6. Practical joking among students | | | 30 | 30 | −37 | | | −43 |
| 7. Likelihood of student demonstration | | | — | | | | | |
| 8. Rejection of introverted students | | −39 | | −30 | | 35 | | |
| 9. Student's interest in marriage and family | 40 | −69 | | −58 | 36 | 38 | | |
| 10. Emphasis on parties and dating | 39 | −50 | | −32 | — | 44 | | |
| 11. Rejection of student who lacks social skills | | | −37 | −37 | 41 | | 32 | 32 |
| 12. Student's conformity in dress | | −45 | | −34 | | 30 | | |
| 13. Administrative pressure on students to conform to rules | | −56 | | −36 | | 34 | | |
| 14. Faculty resentment of students' criticism | 45 | −70 | 37 | | −31 | 56 | −32 | −40 |
| 15. Student's eagerness to argue and debate issues | −33 | 59 | −34 | | | −31 | 41 | |
| 16. Competition for leadership positions | 35 | −61 | | −33 | | 41 | — | |
| 17. Individualism of students | | 61 | | 34 | | −42 | | — |
| 18. Students' interest in competitive sports | 36 | −57 | | | | 50 | | −36 |

TABLE 2—Continued

Correlations[a] of EAT Variables with 18 Ratings of the College Environment and with 21 Perceived Effects of the College
($N = 76$ Colleges, 15–216 student raters per college)

| Item | EAT variable | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Personal orientation | | | | | |
| | Size of student body | Intelligence level of student body | Realistic | Intellectual | Social | Conventional | Enterprising | Artistic |
| Perceived effects | | | | | | | | |
| Improved | | | | | | | | |
| 19. Clarity of vocational plans | | | | | | | | |
| 20. Study habits | −41 | | | | | −37 | | |
| 21. Poise in groups | 34 | −44 | | | | | | |
| 22. General personal development | | | | | | | | |
| Increased | | | | | | | | |
| 23. Interest in art and music | −41 | 47 | −59 | | 51 | −33 | 40 | 59 |
| 24. Knowledge of etiquette, social graces | 41 | −59 | | −42 | | 36 | | |
| 25. Ambition | | −54 | | | | | −34 | |
| 26. Interest in sports | | −33 | 40 | | −42 | | −34 | −47 |
| 27. Concern over grooming and appearance | 57 | −70 | | −41 | | 42 | −31 | |
| 28. Amount of drinking | | | | | | | | |
| 29. Interest in world affairs | | | | | 30 | | | |
| 30. Concern for feelings of others | | | | | | | | 31 |
| 31. Interest in science | 34 | −44 | 53 | | | | −55 | −45 |
| 32. Self-confidence | | −46 | | | | | | |
| 33. Tendency to argue and debate issues | | | | | | | | |
| More positive attitude toward | | | | | | | | |
| 34. Formal religion | | −32 | | | | 32 | | |
| 35. Fraternities and sororities | | −37 | 57 | | −38 | | −51 | −49 |
| 36. Conservative political beliefs | 44 | −51 | 44 | | −37 | 45 | −34 | −49 |
| 37. Marriage | | −44 | | | | | | |
| 38. Number of Assets | −40 | 52 | −68 | | 53 | −30 | 60 | 59 |
| 39. Number of Liabilities | | | | | | | | |

Note.—Predicted relationships which were confirmed are shown in boldface; predicted relationships which were not confirmed are shown by a dash (—).

[a] Only significant ($p < .01$) correlations are shown. Decimal points have been omitted.

EAT variables and the 18 ratings of the college environment shown in Table 2, 81 (56%) were significant ($p<.01$). Thirty-four percent (57 out of 168) of the correlations computed between EAT variables and the 21 perceived "effects" of college shown in Table 2 were significant.

To facilitate interpretation, the specific results for each of the EAT variables are discussed separately.

*Size of the Student Body.* Both of the predicted relationships (Items 1 and 2) were confirmed. As expected, the pattern of correlations involving Size tends to be the opposite of the pattern involving Intelligence.

*Intelligence Level of the Student Body.* All three predicted relationships involving Intelligence (Items 3, 4, and 5) were confirmed. Twenty-eight of the 39 items had significant relationships with Intelligence—more than with any other EAT variable. In fact, 21 of the 33 items which were significantly related to one or more EAT variables had their highest correlation with Intelligence. This result supports previous findings (Astin, 1962; Astin & Holland, 1961) that the ability level of the student body tends to be a major factor in determining the college environment.

The pattern of correlations involving Intelligence suggests that the environments at institutions with student bodies of relatively high ability are characterized by an emphasis on verbal aggression (Item 15), scholarship (Items 3, 4, and 9), and individualism (Items 12 and 17); by little emphasis on social activities (Items 1, 10, and 16) and sports (Item 18); and by a relatively low degree of administrative control (Items 13 and 14). Intelligence is also related to a perceived increase of interest in art and music and a perceived decrease of interest in science (Items 23 and 31). Other findings which may have significance are the negative relationships between Intelligence Level of the Student Body and the perceived effect of the college experience on "self-confidence" and "ambition." These latter relationships suggest the interesting hypothesis that continued exposure to a peer group of unusually high intelligence has a detrimental effect on the self-confidence and ambition of some students.

*Realistic Orientation.* Results with the two "acting-out" items lend only equivocal support to this hypothesized aspect of the Realistic environment. Only one of the items ("practical joking") was significantly related in the predicted direction, and the correlation was low ($r=.30$). Students at colleges with high Realistic Orientations tend to report fewer assets of the college on the open-ended question ($r=-.68$), but they do not at the same time report more liabilities. They say that college has decreased their interest in art and music and increased their interest in science. The Realistic Orientation had a higher correlation ($r=.57$) with "positive attitudes toward fraternities and sororities" than any other EAT variable.

*Intellectual Orientation.* Results with both items for the Intellectual Orientation confirmed the predictions (Items 8 and 9). All items pertaining to the importance of social groups and interpersonal skills (2, 10, 11, 12, 16, 17, 24, and 27) tended to be negatively related to the Intellectual Orientation.

*Social Orientation.* Only one of the two items (11) produced significant results in the predicted direction for the Social Orientation, although the correlation of that item with the Social Orientation ($r=.41$) was higher than with any other EAT variable. The Social Orientation was also the only EAT variable which was significantly related ($r=.30$) to the perceived effect of college on "interest in world affairs" (Item 29).

*Conventional Orientation.* All three items constructed for the Conventional Orientation (Items 12, 13, and 14) produced significant results in the predicted direction. Other relatively high correlates of this Orientation included the "time spent by students in the library" (r=−.50), the estimated "snob" appeal of the social groups ($r$=.44), the estimated quality of teaching ($r$=−.59), student interest in competitive sports ($r$=.50), a perceived increase in concern over grooming and personal appearance ($r$=.42), and more conservative political beliefs ($r$=.45).

*Enterprising Orientation.* Only one of the two items designed for the Enterprising Orientation (Item 15) produced significant results: "Students are eager to argue and debate issues" ($r$=.41). The Enterprising Orientation also had the largest correlation of all the EAT variables ($r$=−.55) with increased interest in science (Item 31).

*Artistic Orientation.* Only one of the two items designed for the Artistic Orientation (Item 18) yielded a significant correlation in the predicted direction ($r$=−.36). The highest correlates of the Artistic Orientation included the estimated "time spent by students in the library" ($r$=.59), the perceived effect of the college on "interest in art and music" ($r$=.59), and the number of assets of the college reported on the open-ended question ($r$=.59).

## DISCUSSION

In general, the results seem to support the construct validity of EAT as a method for assessing the college environment. Fourteen of the 18 predicted relationships were confirmed, even though the items were selected to test relatively remote hypotheses about each EAT variable. Furthermore, there were many substantial correlations between the EAT variables and the effects of the college as perceived by the student. However, only partial support was obtained for some of the proposed interpretations of the Realistic, Social, Enterprising, and Artistic Orientations.

Some of the findings in the present study seem to be relevant to a recent study of differential college effects (Astin, 1963), in which it was found that the aspiration of the talented student to obtain the PhD is negatively affected both by the Size of the Student Body and by the Conventional Orientation. In the present study, Size and the Conventional Orientation had similar patterns of relationships to the validation items. For example, institutions with high scores on these two EAT variables tend to emphasize sports and social activities (Table 2, Items 2, 10, 16, 24, and 27) at the expense of academic pursuits and scholarship (Items 3, 4, and 9). According to the student informants, faculty members at these institutions have fewer contacts with the students (Item 1), are poorer teachers (Item 5), and are resentful of criticism (Item 14). In addition, Size and the Conventional Orientation were the only EAT variables significantly related to the perceived effect of college on "improvement in study habits" (Item 20; for both variables the correlation was negative). Thus, it seems plausible that a college environment characterized by this complex of attributes tends to discourage the pursuit of higher academic degrees.

The large proportion of variance in the 39 validation items which was attributable to Intelligence Level of the Student Body poses a problem. This finding can be explained partially by the spuriously high correlation between Intelligence and the other EAT variables in the present sample of 76 institutions; in the larger sample of 335 institutions (Table 1), Intelligence accounts for only a small proportion of variance in the other EAT measures. It is clear that more precise differentiation among the eight EAT variables will require additional studies of less biased samples of institutions.

## REFERENCES

ASTIN, A. W. An empirical characterization of higher educational institutions. *J. educ. Psychol.*, 1962, **53**, 224–235.

ASTIN, A. W. Differential college effects on the motivation of talented students to obtain the PhD. *J. educ. Psyckol.*, 1963, **54**, 63–71.

ASTIN, A. W., & HOLLAND, J. L. The Environmental Assessment Technique: A way to measure college environments. *J. educ. Psycol.*, 1961, **52**, 308–316.

HOLLAND, J. L. Some explorations of a theory of vocational choice: One- and two-year longitudinal studies. *Psychol. Monogr.*, 1962, **76**(26, Whole No. 545).

PACE, C. R., & STERN, G. G. An approach to the measurement of psychological characteristics of college environments. *J. educ. Psychol.*, 1958, **49**, 269–277.

(Received February 8, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 4, 227–230

# RELATIONSHIP OF REACTIVE INHIBITION TO READING SKILL ATTAINMENT

WAYNE OTTO
*University of Oregon*

AND

ROBERT C. FREDRICKS
*Portland Continuation Center*

The purpose of this study was to examine the relationship of reactive inhibition to skill attainment in reading, and the details of measuring reactive inhibition by an inverted number printing task. 100 Ss—equal numbers of good and poor readers in Grades 4, 5, 6, 7, and 8—served in the study. A repeated measures analysis of variance was employed. It was found that as expected, poor readers generate reactive inhibition more rapidly than good readers, and the inverted number printing task appears to be satisfactory for quantifying reactive inhibition in children.

A recent study by Lynn (1960) examines the effect of reactive inhibition upon the attainment of skill in reading. The data reveal a low but significant ($-.23$, $p < .05$) negative correlation between the accumulation of reactive inhibition, as measured by an inverted number printing task, and ability in reading. The finding gives some support to the hypothesis, based upon Hullian theory, that the more rapidly reactive inhibition accumulates the less adequately certain skills are acquired.

Lynn's study is rather unique in that the subjects were children between the ages of 8 and 11 and inverted number printing was employed to measure reactive inhibition. The traditional approach has been to employ inverted alphabet printing or rotary pursuit tasks with adult subjects in studies involving reactive inhibition. An inverted number printing task appears to have a distinct advantage (see below) for use with children. However, Lynn has not reported in detail upon the implementation of the number printing task.

The present study has two concerns. (*a*) The relationship of reactive inhibition and the attainment of reading skill is examined further. The hypothesis is that good readers generate less reactive inhibition than poor readers in a given period of time. A demonstration of a clear relationship between reactive inhibition and skill attainment in reading would clear the way for studies designed to get at implications for instructional procedures. (*b*) The details of measuring reactive inhibition by inverted number printing are examined and reported. This appears to be a necessary prelude to further use of what appears to be a useful method of quantifying reactive inhibition in children.

## METHOD

### *Measurement of Reactive Inhibition*

Reactive inhibition was measured by an inverted number printing task. The procedure generally followed that employed by Archer and Bourne (1956), although they used inverted alphabet printing. Briefly, the subjects were told about the nature of the task and that the digit 6 is 9 upside down and vice versa. The latter was done to reduce variability. The subjects were told to print the numbers from 1 to 10 as quickly as possible from left to right starting in the upper left-hand corner, placing one number in each .5-inch square on the printing sheet. They were to print right over errors and not to attempt to erase.

The experimenter put the inverted numbers on a chalk board after the subjects had attempted to make the inversions on their own. This was found to be desirable in order to get uniform inverted number forms from the subjects. Once this was done, the children had little difficulty; and they seldom consulted the chalk board, despite the fact that the experimenter's numbers were left there. The inverted numbers appear to have a distinct advantage over inverted letters when subjects are children. In pilot tryouts, it was found that poor readers in particular had extreme difficulty in making inverted letters, even after coaching had been done.

Subjects were told to skip a square and continue with the next number when told to "skip." This signal was given every 30 seconds, i.e., at the end of each trial. After 12 massed trials they were told to stop, and a 5-minute rest followed. The rest period was spent with the experimenters chatting with the subjects. Ten seconds before the end of the rest period, the subjects were asked to "get ready," make 5 Xs after the last number on their papers, and resume printing when told to "begin." Four 30-second massed trials followed the rest interval.

Typically, when last prerest and first postrest trials are compared, there is a marked gain in the number of digits produced after the rest interval, the gain being the phenomenon of reminiscence. It is assumed that the amount of reminiscence reflects the amount of reactive inhibition generated, the argument being that reminiscence is due to the dissipation of reactive inhibition. See Kimble (1949) for a more extended discussion of reminiscence and reactive inhibition in tasks similar to the present.

## *Subjects and Design*

Both the reading level and grade placement of subjects are considered in the present study: subjects are good and poor readers in Grades 4, 5, 6, 7, and 8. It was felt that the grade level (chronological age) of subjects would be a significant determiner of digits produced. Therefore, a repeated measures factorial design (Edwards, 1950) was employed: two levels of reading ability, five grade levels, and prerest and postrest learning trials.

A total of 100 children from four elementary schools in a suburban school district served in the experiment. All subjects had IQs in the average range (90–115), according to test scores corroborated by the classroom teachers' judgments. Good readers were children who, according to test scores and teacher evaluations, had reading abilities above the seventieth percentile on grade level norms. Poor readers were children who were certified for the remedial reading program. To be certified, a child must have at least average intelligence (individual IQ test score of 90 or above) and reading achievement at least 1 year below grade level in the fourth grade and 2 or more years below in Grade 5 and above.

Ten good readers and 10 poor readers at each grade level—4, 5, 6, 7, and 8—provided data for the study. In all, 114 children completed the printing task, but 14 data sheets were discarded because the blank spaces that set apart trials had been omitted.

## *Procedure*

Subjects were run in groups of four to eight. Two experimenters were present at each session, which expedited accurate timing of trials and supervision of the number printing activity. All subjects were run during a 3-week period. When a group had completed the task, they were asked not to discuss it with their friends until the experimenters had left the school. This procedure appears to have successfully precluded advance knowledge of the task and resultant practice.

A study by Wasserman (1951) shows that degree of motivation has a significant effect upon inhibitory potential. An effort was made, therefore, to secure reasonable motivation. The subjects were told that their performance in making "upside-down numbers" would tell the experimenters "something about how people learn," but they were assured that the task was not a test. This procedure, combined with informal chatting to secure rapport, appeared to be successful in securing motivation without anxiety.

## RESULTS AND DISCUSSION

The mean number of inverted numbers printed in Trial 1 was exactly the same for good and poor readers. This is shown in Table 1, which presents mean responses on selected prerest and postrest trials. Inspection of Table 1 also shows that a plotted acquisition curve would show limited positive acceleration for the good readers and negligible acceleration for the poor readers. The implication appears to be that whatever skill was developed by the poor readers was nullified by accumulating reactive inhibition. The postrest results are classic, in that on Trial 13 good and poor readers again performed almost identically; by the final trial (Trial 16), the poor readers' performance was again beginning to diverge below that of the good readers.

TABLE 1

MEAN RESPONSES ON SELECTED TRIALS

| Reading level | Trial |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|---|
|  | 1 | 5 | 8 | 11 | 12 | Rest | 13 | 16 |
| Good | 9.98 | 10.84 | 11.44 | 11.48 | 11.5 | — | 14.02 | 13.54 |
| Poor | 9.98 | 10.28 | 10.64 | 10.38 | 10.1 | — | 14.08 | 13.02 |

When the prerest and the postrest trials were considered separately, it was found that there were no significant between-trial differences in digits produced (i.e., Trial 1 did not differ from Trial 2; Trial 2 did not differ from Trial 3, etc.). Therefore, only last prerest (Trial 12) and first postrest (Trial 13) trial results are considered in subsequent analyses.

Mean responses (digits produced) on Trials 12 and 13 by good and by poor readers at each grade level are presented in Table 2. A repeated measures analysis of variance (Edwards, 1950) of digits produced on Trials 12 and 13 is summarized in Table 3. There is significant grade level difference ($p<.01$)—the higher the grade level, the greater the number of digits produced tends to be (see Table 2). As expected, there is a highly significant between-trial difference ($p<.001$); within the present theoretical framework, the gain from prerest to postrest performance (see Table 2) demonstrates the dissipation of reactive inhibition during rest.

There is not a significant difference between reading levels; that is, when overall performance is considered good readers do not differ from poor readers. However, poor readers do differ from good readers in the gain in digits produced after rest. This finding is indicated by the significant Trial × Reading Level interaction ($p < .01$) and demonstrated in Table 2. If the prerest and postrest performances of good and poor readers were plotted, the slopes would differ significantly; that is, the performance of the poor readers increased more sharply than that of the good readers. The implication is that the difference in rate of increase was due to the dissipation of more reactive inhibition by the poor readers. The hypothesis

TABLE 2

MEAN RESPONSES ON LAST PREREST AND FIRST POSTREST TRIALS

| Grade | Trial 12 |  |  | Trial 13 |  |  |
|---|---|---|---|---|---|---|
|  | Good | Poor | All subjects | Good | Poor | All subjects |
| 4 | 8.1 | 7.3 | 7.7 | 8.9 | 10.4 | 9.7 |
| 5 | 8.7 | 10.3 | 9.5 | 11.1 | 14.5 | 12.8 |
| 6 | 13.5 | 9.4 | 11.2 | 16.9 | 14.0 | 15.5 |
| 7 | 12.8 | 9.8 | 11.3 | 15.3 | 15.0 | 15.2 |
| 8 | 14.4 | 13.7 | 14.1 | 17.9 | 16.5 | 17.2 |
| All grades | 11.5 | 10.1 | 10.8 | 14.0 | 14.1 | 14.1 |

TABLE 3

ANALYSIS OF VARIANCE OF LAST PREREST AND FIRST POSTREST TRIALS

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Grade Level (G) | 4 | 276.1875 | 11.425* |
| Reading Level (R) | 1 | 22.4450 | — |
| G × R | 4 | 50.5113 | 2.099 |
| Subjects/Groups | 90 | 24.0683 | 4.128* |
| Trials (T) | 1 | 528.1250 | 90.578* |
| T × G | 4 | 6.5625 | 1.126 |
| T × R | 1 | 26.6450 | 4.570** |
| T × G × R | 4 | 4.3263 | — |
| Subjects × T/Groups | 90 | 5.8306 |  |

* $p < .001$.
** $p < .01$.

that good readers generate less reactive inhibition than poor readers in a given period of time is, therefore, accepted on the basis of the present data.

The results of the present study support Lynn's (1960) finding that reactive inhibition appears to be associated with reading attainment. This may, in turn, lend support to the notion that, at least in the early stages, certain school learning (i.e., simple addition and reading) takes place by instrumental conditioning (Peel, 1956). Whatever the theoretical framework, the relationship demonstrated by the present study appears to hold promise for further investigations. It would be interesting, for example, to substitute good and poor achievers in, say, arithmetic, spelling, or handwriting for the good and poor readers of the present study. If this were done, we would be better able to see the role of reactive inhibition in learning disability problems. Once this role is understood, it will be possible to design studies that have implications for remedial teaching.

The present study also demonstrates the feasibility of using an inverted number printing task to measure reactive inhibition in children. In addition, the need to control the grade level (chronological age) of subjects in future studies is made clear. The details of implementing the number printing task are presented in the hope that the present task will be a useful tool in future research studies.

## REFERENCES

ARCHER, E. J., & BOURNE, L. E., JR. Inverted-alphabet printing as a function of intertrial rest and sex. *J. exp. Psychol.*, 1956, **52**, 322–328.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Holt, Rinehart & Winston, 1950.

KIMBLE, G. A. Performance and reminiscence in motor learning as a function of the degree of distribution of practice. *J. exp. Psychol.*, 1949, **39**, 500–510.

LYNN, R. Individual differences in introversion-extraversion, reactive inhibition, and reading attainment. *J. educ. Psychol.*, 1960, **51**, 318–321.

PEEL, E. A. *The psychological basis of education.* Edinburgh: Oliver & Boyd, 1956.

WASSERMAN, H. N. The effect of motivation and amount of pre-rest practice upon inhibitory potential in motor learning. *J. exp. Psychol.*, 1951, **42**, 162–172.

(Received March 25, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 5, 231-237

# A SEX DIFFERENCE IN THE AGE OF ONSET OF THE PARENT-CHILD RESEMBLANCE IN INTELLIGENCE[1]

MARJORIE P. HONZIK
*Institute of Human Development*
*University of California, Berkeley*

In a representative urban sample of 248 children, years of schooling of both parents showed an increasing relationship to their son's and daughter's IQs between 21 mo. and 15 yr. The parent-child correlation was observable and became statistically significant for girls by 3 yr. but not until 5 yr. for boys. Earlier parent-daughter resemblance was also noted when ratings of mother's intelligence and social status were correlated with the children's scores. Siblings' IQs yielded comparable sex differences in relation to parents' education. This finding was further confirmed in the correlations of the IQs of a 2nd generation sample with the IQs of the parent in the original group at age 15 yr. A sex difference in rate of mental growth is suggested as the major determinant of these results.

Sex differences in mental test performance are not often significant when the average scores of boys and girls on standard tests are compared, but hidden in these averages may be changing rates of mental growth which are of considerable interest to all who are charged with the education of children. In the present study the concern is with sex differences in the rate of increase in the parent-child resemblance in ability occurring during childhood.

In a first study of parent-child resemblance in intelligence, Jones (1928) reported a correlation of .50 between the intelligence test scores of parents and their children. More recently, reports from longitudinal investigations have shown that little or no resemblance exists between the abilities of parents and of their infants, but that a relationship develops late in the second, or in the third, year of life and increases during childhood until a correlation ranging from .25 to .75, with an average of .50, is reached (Bayley & Jones 1937; Honzik, 1940, 1957; Kagan & Moss, 1959). It was thought at first that this increasing resemblance was the result of environmental influence. However, Skodak and Skeels' (1949) longitudinal study of adopted children reported a similar age change between the abilities of adopted children and their *true* parents who placed the children for adoption during the first months of life, but no significant relationship between these same adopted children's IQs and the abilities of their adopting parents. This significant finding points to a genetic basis for at least some of this relationship.

The question to which this paper is addressed is the extent to which there is a sex difference in these relationships. Higher mother-child than father-child coefficients would suggest the importance of environmental stimulation; higher same-sex correlations might be the result of genetic factors or differential stimulation of boys and girls by their parents. Higher opposite-sex correlations might reflect sex linkage or environmental effects. Differences in the age at which boys' and girls' IQs correlate significantly with measures of parental ability would suggest a sex difference in the rate of mental growth.

---

[1] This study was supported in part by a grant from the Ford Foundation and a United States Public Health Service grant, No. MH-06238-02.

Increasing parent-child similarities in ability are reported for a California sample by Bayley (1954), an Ohio sample by Kagan and Moss (1959), and a London sample by Hindley (1962). In these three investigations results are given for boys and girls separately, permitting an analysis of sex differences. Bayley (1954) concluded for the 28 boys and 28 girls in the Berkeley Growth Study:

> there appears to be little sex difference in the parent-child relationships in mental ability, in this group, although the girls' scores are oftener than the boys' significantly correlated with parental status [p. 6].

Kagan and Moss (1959) reported, for 59 boys and 40 girls, product-moment correlations between the child's Stanford-Binet IQ at 3, 6, and 10 years and (*a*) parental education, and (*b*) parental Otis IQ. In this study maternal education (years of schooling) predicts IQs of both sons and daughters better than paternal education at all three ages. Additionally, maternal education was a better predictor of the girl's IQ than the boy's. The correlations between the child's IQ and the education of both parents were statistically significant at 3 years for the girls but not until 6 years for the boys.

In the British study Hindley reports a significant sex difference in the relation of the child's test score to father's occupation for 40 boys and 37 girls tested on the Griffiths scales at 6 months and 1½ years, and the Stanford-Binet at 3 and 5 years. For the girls, a relationship between social class and child's IQ is noted at 1½ years which becomes statistically significant at age 3 years ($p < .001$); but for boys, the relationship only begins to appear at 5 years ($p < .10$). This finding is in line with Kagan's report of higher correlations between the daughter's IQ and parental education, especially at 3 years. However, Kagan also presents results using the Otis IQ as the parental ability measure showing a slightly higher father-son than father-daughter correlation in IQ at 3 years which does not fit this trend of an earlier appearance of the parent-child similarity in ability in girls than in boys.

In summary, two of these three longitudinal studies report sex differences in the parent-child relationship in ability (whether the parent's ability is measured by occupational status, years of schooling, or IQ) which are interesting but not entirely consistent and which will be of interest to explore in still another developmental study—the Guidance Study.

## SAMPLE AND PROCEDURE

The Guidance Study sample of 248 children was selected from the Birth Certificate Registry and is a representative sample of the children born in Berkeley during an 18-month period (from January 1, 1928, to June 30, 1929). This urban sample was divided into two matched groups of 124 children on the basis of socioeconomic factors before the children came to the institute for their first mental test at 21 months. One of these groups is referred to as the Guidance group because of the program of intensive interviewing of parents and children. The second, or Control group, was measured and tested at comparable age levels but there were no intensive interviews until the children were 17 years old (Macfarlane, 1938). In addition to the 21-month test, mental tests were administered to the children in both groups at ages 3, 3½, 4, and 5 years; at yearly intervals until 10 years; at 12 or 13 years; and at 14 or 15 years. The children were tested on the California Preschool scales during the first 5 years (Honzik, 1938), the 1916 revision of the Stanford-Binet at 6 and 7 years, and Form L or M of the 1937 Stanford Revision at subsequent ages.

## RESULTS FOR THE GUIDANCE STUDY SUBJECTS

Correlations between the education of the mother (number of years of schooling) and the IQs of sons and daughters are presented for successive ages in Table 1. For over 100 cases a slight mother-daughter correlation is noticeable at 21 months which becomes statistically significant at age 3½ years. In contrast, with an equally sizable $N$, the mother-son correlation is positive at age 3½ years ($r = .24$), but does not become significant at the .01 level until 5 years. Similarly, the correlation between the father's education and his daughter's IQ is significant at 3 ($r = .32$) but it is not until 5 years that the boy's IQ is significantly related to his father's education. This sex difference in the age level at which the parent-child resemblance becomes significant is clear. At ages 21 months, 3 years, and 4 years, these sex differences are statistically significant at the .05 level or better. A further finding (see Table 1) is the increase in parent-son resemblance until age 14–15 years. And this is true for the father-*son* as well as the mother-*son* coefficients. This fact suggests either (*a*) continued environmental influence or (*b*) maturing of mental functions in boys until middle adolescence. Consideration of the Skodak-Skeels results points to the latter interpretation.

Table 1 gives the results for the combined Guidance and Control groups. Correlations were also computed for these two subgroups, each of which is comparable in size with those used in previous longitudinal studies reporting sex differences. In both subsamples, Guidance and Control groups, the parent-child correlations became positive and statistically significant at an earlier age for girls than for boys.

Since both the mother's and father's education was significantly related to the daughter's IQ before the son's, it was of interest to determine the combined effect of the father's and mother's contribution by computing a multiple $R$ between the boy's and girl's scores and their parents' education (see Table 1). In this sample the number of years of schooling of the mothers and fathers correlates to the extent of an $r$ of .74 (Atherton, 1962). This $r$ of .74 is suf-

TABLE 1

CORRELATIONS BETWEEN THE CHILD'S IQ AND PARENTAL EDUCATION FOR GUIDANCE STUDY SUBJECTS

| Age in years | $N$ cases | | Mother's Education × Child's IQ | | Father's Education × Child's IQ | | Parental Education × Child's IQ | |
|---|---|---|---|---|---|---|---|---|
| | Boys | Girls | Boys' $r$ | Girls' $r$ | Boys' r | Girls' r | Boys' $R$ | Girls' $R$ |
| 1¾ | 109 | 114 | .00 | .13 | .01 | .17 | .00 | .17 |
| 3 | 110 | 113 | .00 | .18 | −.08 | .32** | .12 | .33** |
| 3½ | 105 | 109 | .24 | .30** | .15 | .32** | .24 | .33** |
| 4 | 105 | 107 | .17 | .30** | .10 | .36** | .17 | .36** |
| 5 | 101 | 109 | .40** | .29** | .23 | .34** | .41** | .34** |
| 6 | 102 | 112 | .25 | .38** | .17 | .43** | .25 | .44** |
| 7 | 101 | 107 | .34** | .34** | .34** | .44** | .36** | .44** |
| 8 | 94 | 104 | .37** | .30** | .27** | .41** | .37** | .41** |
| 9 | 93 | 102 | .41** | .37** | .29** | .41** | .41** | .42** |
| 10 | 95 | 102 | .31** | .36** | .27** | .40** | .32** | .41** |
| 12 or 13 | 92 | 100 | .42** | .34** | .35** | .42** | .42** | .42** |
| 14 or 15 | 83 | 85 | .44** | .25 | .41** | .33** | .46** | .33** |

** $p < .01$.

ficiently high so that the multiple *R*s are not markedly higher than the zero-order *r*'s. However, the trend is clear. The parent-daughter resemblance ($R = .33$) is significant at the .01 level at 3 years, but the parent-son correlations do not become significant to this degree until age 5.

Parents' education, or years of schooling, is at best a rough measure of their ability. A second approximation, or indirect measure of paternal ability, is Warner's (Warner, Meeker, & Eels, 1949) index of status characteristics of the family at the birth of the child. This index which reflects paternal intelligence (since it is based on: fathers' occupation, source of income, house type, and dwelling area), correlates .64 with fathers' and .62 with mothers' education (Atherton, 1962). The daughter's IQ is significantly correlated with this index at age 3 ($r = .27$), but it is not until 5 years that this relation is significant at the .01 level in boys ($r = .37$). At 3 and 4 years the sex differences in the coefficients are significant at the .05 and .07 levels, respectively (see Figure 1). This finding is very similar to that reported by Hindley, who found a highly significant relationship between the father's occupational rating and his daughter's IQ at 3 years ($p < .001$) but a significance level of .10 was not reached for the sons until 5 years.

Estimates of the mother's intelligence are still another measure of maternal ability which have been related to the children's test performance. The intelligence of the mothers of the subjects in the Guidance group was rated by two members of the staff on a 7-point scale after a series of interviews and observations of the mother and her child beginning at 21 months (Macfarlane, 1938). The interrater agreement for these ratings was .68. This rating correlated .73 with the mother's education. The average of the two ratings has been correlated with the boy's and girl's mental test scores at successive age levels (see Figure 2). The trend observed for the other indexes of parental ability is seen again, namely, that the mother-child re-

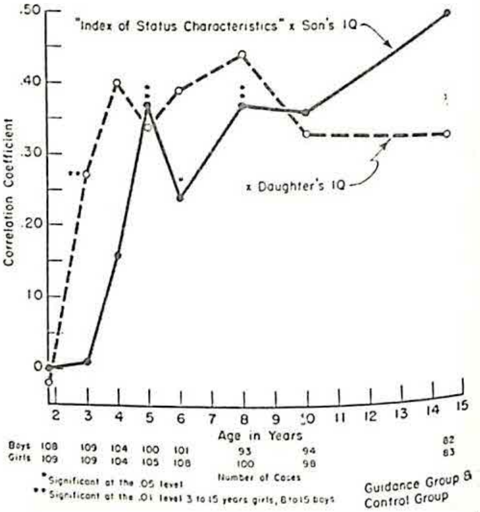


FIG. 1. Age changes in the relationship of Warner's "index of status characteristics" to the IQs of the sons and daughters at successive ages. (Combined Guidance and Control groups.)

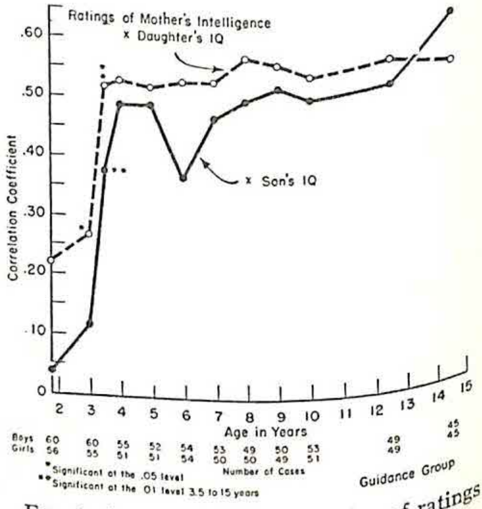


FIG. 2. Age changes in the relation of ratings of mother's intelligence to the son's and daughter's IQs at successive ages. (Guidance group only.)

semblance in "intelligence" occurs at an earlier age for girls than for boys.

Two rough and indirect measures of fathers' mental ability (education and Warner's index of status characteristics) and two of mothers' (education and ratings of intelligence) show a significant correlation with the tested intelligence of daughters by the age of 3 years but not until 4 or 5 years for their sons. This finding, which is in accord with that reported by Hindley (1962) and in part by Kagan and Moss (1959) in two other longitudinal samples, suggests that individual differences in mental ability become stabilized earlier in girls than in boys. This sex difference is not large but is statistically significant and it appears in spite of the use of gross approximations of parental ability and scores on mental tests which were designed to be equally difficult for boys and girls (cf. Terman & Merrill, 1937, p. 34).

Although the results for the Guidance Study subjects are significant, they are not in exact agreement with some of the findings of Kagan and Bayley so that validation in other samples of parents and children should be considered. Data are available for two more groups in the Guidance Study. These are the (*a*) siblings and (*b*) children of the Guidance Study subjects. These additional samples are large but were not so carefully selected to be representative of an urban population as were the Guidance Study subjects. This is especially true for the sibling group which in some instances includes more than one sibling in a family and where of the 205 tests considered, 67 were retests. Thus the sibling sample is not a longitudinal sample, nor is the second generation sampling of Guidance Study subjects' children who ranged in age from 1 to 10 years when they came in for a single mental test.

## RESULTS FOR THE SIBLINGS OF THE GUIDANCE STUDY SUBJECTS

The subjects' siblings were examined on the same tests as the subjects during the years that the subjects were coming in for tests, measurements, and interviews at yearly and half-yearly intervals. Table 2 presents the correlations between the sibling's IQ and the mother's and father's education. These coefficients are based on the same parental measures as for the subject but

TABLE 2

CORRELATIONS BETWEEN THE SIBLINGS' MENTAL TEST SCORES[a] AND THEIR MOTHERS' AND FATHERS' EDUCATION

| Age range in years | Mothers' Education × Mental Test Scores of: | | | | Fathers' Education × Mental Test Scores of: | | | |
|---|---|---|---|---|---|---|---|---|
| | Brothers | | Sisters | | Brothers | | Sisters | |
| | *N* | *r* | *N* | *r* | *N* | *r* | *N* | *r* |
| 1¾–3½ | 16 | .12 | 21 | .69** | 16 | .19 | 21 | .28 |
| 4–5 | 26 | .05 | 27 | .53** | 26 | .09 | 27 | .76** |
| 6–8 | 33 | .20 | 20 | .54* | 34 | .35* | 20 | .51* |
| 9–15 | 27 | .34 | 17 | .61** | 27 | .39* | 17 | .79** |

[a] The tests used were those given to the subjects at the same age level, either the California Preschool scale or the Stanford Revision, Form L.

* $p < .05$ or better.

** $p < .01$ or better.

the parental measures are now considered in relation to the test scores of a subject's brother or sister. An approximately equal number of siblings of the same and opposite sex as the subject was included. The results (see Table 2) are similar to those found for the Guidance Study subjects (see Table 1) and strengthen the probability that parent-child resemblance in ability occurs earlier in girls than boys. The correlations between the brother's IQ and his parent's years of schooling show a gradual increase with age. These correlations become significant at the .05 level at age 6 years and after. In this sibling sample the girl's (sister's) IQ is highly correlated with parental education during the preschool years. This finding is in line with the coefficients reported by Kagan and Moss.

## RESULTS FOR THE SECOND GENERATION CHILDREN OF THE GUIDANCE STUDY SUBJECTS

The children of the Guidance Study subjects (a second generation sample) were included in a testing program when their parents (the Guidance subjects) were seen in a follow-up study at 30 years (between 1958–61). The same mental tests were administered to the second generation as to their parents at the same age. Thus, a subject's children aged 3 and 6 years at the time of the follow up would be given, respectively, the California Preschool scale and the Stanford Revision. The subjects, or first generation sample, were not retested in the follow up. The latest age level at which the Stanford Revision was administered to the subjects was at 14 or 15 years so the subjects' IQs at this age were correlated with the scores of their sons and daughters (second generation) at whatever age the children had attained when their parents were 30 years old. The correlations between the IQs of the subjects' children and those of the subjects are shown in Table 3.

For the second generation of children, the parent-child correlations are higher for the girls than the boys at 3 and 4 years, but at 5 years significantly higher correlations are obtained for the boys. These findings agree with the results reported for the first generation study but are not as consistent or significant.

## DISCUSSION

A question asked at the beginning of this paper was the extent to which the mother-child resemblance is greater than the father-child resemblance in intelligence. Bayley and Kagan both reported higher mother-child than father-child relationships which led Kagan to conclude, "environmental pressures may play an important role in the determining of the child's IQ score." In the present study, for a much larger sample, paternal education correlated higher than maternal education with the daughter's IQ at all ages. Higher cross-sex correlations were also obtained between the mother's education and the son's IQ. This finding was consistent in two subsamples (Guidance and Control

TABLE 3

CORRELATIONS BETWEEN THE 14–15 YEAR IQS OF THE SUBJECTS AND THE MENTAL TEST SCORES OF THEIR CHILDREN

| Children's age in years | Subjects' IQs at 14–15 Years ×: | | | |
|---|---|---|---|---|
| | Son's IQ | | Daughter's IQ | |
| | *N* | *r* | *N* | *r* |
| 2 | 20 | .16 | 21 | .15 |
| 3 | 28 | .06 | 24 | .23 |
| 4 | 23 | .33 | 19 | .48* |
| 5 | 26 | .52** | 21 | .21 |
| 6 | 15 | .55* | 17 | .39 |
| 7–10 | 33 | .47** | 22 | .28 |

* $p \leq .05$ or better.
** $p < .01$ or better.

groups). The results for the siblings of the Guidance Study subjects did not clarify these discrepancies since the relationships in this sample were unusually high for the sisters. The higher opposite-sex relationships obtained for the Guidance Study subjects is an interesting finding but one which has not been found in other samples.

The earlier appearance of a parent-child resemblance in ability in girls is evident although the sex difference is not so great as to preclude instances of reversals in this trend. One explanation for this finding may be that the preschool-aged girl is more cooperative and controlled in the test situation than the boy in this age period, but this greater cooperation and control may be a reflection of generally earlier maturing of cognitive function in girls. It is probable that the maturing of the central nervous system proceeds more slowly in the male than the female as is true for other aspects of physical growth (e.g., height). This slight sex difference in maturation rate might then account for the sex difference in the age at which the parent-child resemblance in intelligence becomes visible. There is other evidence for a sex difference in the pattern of mental growth. Bradway and Thompson (1962) report that males show a significantly greater gain in IQ ($p < .01$) after adolescence than do females. Results from this and other studies suggest that it is only in longitudinal studies that sex differences in mental growth can be seen in their full perspective. In other words, when sex differences are slight or related to slowly maturing functions, it may be only over a prolonged time period that they become noticeable to a significant degree.

## REFERENCES

ATHERTON, KARYL M. A comparison of solutions obtained in factor analyses of socio-economic variables. *Psychol. Rep.*, 1962, **11**, 259–273.

BAYLEY, NANCY. Some increasing parent-child similarities during the growth of children. *J. educ. Psychol.*, 1954, **45**, 1–21.

BAYLEY, NANCY, & JONES, H. E. Environmental correlates of mental and motor development: A cumulative study from infancy to six years. *Child Develpm.*, 1937, **4**, 329–341.

BRADWAY, KATHERINE P., & THOMPSON, CLARE W. Intelligence at adulthood: A twenty-five year follow-up. *J. educ. Psychol.*, 1962, **49**, 1–14.

HINDLEY, C. B. Social class influences on the development of ability in the first five years. In G. Nielson (Ed.), *Child and education: Proceedings of the XIV International Congress of Applied Psychology*. Vol. 3. Copenhagen: Munksgaard, 1962. Pp. 29–41.

HONZIK, MARJORIE P. The constancy of mental test performance during the preschool period. *J. genet. Psychol.*, 1938, **52**, 285–302.

HONZIK, MARJORIE P. Age changes in the relationship between certain environmental variables and children's intelligence. *39th Yearb. Nat. Soc. Stud. Educ.*, 1940, **39**, 185–205.

HONZIK, MARJORIE P. Developmental studies of parent-child resemblance in intelligence. *Child Develpm.*, 1957, **28**, 215–228.

HONZIK, MARJORIE P. Fluctuations in mental test scores. In, *Physical and behavioral growth: Report of 26th Ross Pediatrics Research Conference*. Columbus, Ohio: Ross Laboratories, 1958. Pp. 54–59.

JONES, H. E. A first study of parent-child resemblance in intelligence. *27th Yearb. Stud Educ.*, 1928, **29**, 61–72.

KAGAN, J., & MOSS, H. A. Parental correlates of child's IQ and height: A cross-validation of the Berkeley Growth Study results. *Child Develpm.*, 1959, **30**, 325–332.

MACFARLANE, JEAN W. Studies in child guidance: I. Methodology of data collection and organization. *Monogr. Soc. Res. Child Develpm.*, 1938, **3**, 1–254.

SKODAK, MARIE, & SKEELS, H. M. A final follow-up study of one hundred adopted children. *J. genet. Psychol.*, 1949, **75**, 85–125.

TERMAN, L. L., & MERRILL, MAUD A. *Measuring intelligence*. Boston: Houghton Mifflin, 1937.

WARNER, E. L., MEEKER, M., & EELS, L. *Social class in America*. Chicago: Science Research Associates, 1949.

(Received June 10, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 5, 238–248

# SOME CORRELATES OF CLASSROOM AND OUT-OF-CLASS ACHIEVEMENT IN GIFTED SCIENCE STUDENTS[1]

EDWIN A. LOCKE
*Cornell University*

The 47 variables measuring intellectual abilities, personality traits, personal habits, and demographic characteristics were intercorrelated for a group of 112 gifted science students and developed into 11 factor index scores. These were correlated with classroom and out-of-class achievement factors developed in an earlier study. Vocabulary and self-control were found to be most highly related to classroom achievement, and school and city size and originality most highly related to out-of-class achievement. There were considerable differences between sexes in the obtained relationships.

A previous paper by Locke (1963) described the development of achievement criteria for a select group of high school science students. Two orthogonal criterion factors were isolated. The first factor contained course grades and summary teacher ratings each in two different situations and will be referred to as "classroom achievement," as it involves accomplishments in assigned school work and classroom behavior. The second factor contained two indexes of the amount of independent work done, one measure of the quality of this work, and one measure obtained from teachers' actual comments about the students. This factor will be referred to as "out-of-class achievement," as it involves accomplishments unrelated to conventional school requirements.[2]

This paper reports the results of a testing program aimed at finding the correlates of these two achievement factors. The types of variables used included intellectual, personality, demographic, and personal habit measures. The paper also has some methodological implications which are brought out in the course of the discussion of results.

## METHOD

### *Subjects*

The subjects tested consisted of 122 high school juniors and seniors (73 boys and 49 girls) attending a Cornell summer school advanced science training program under the auspices of the National Science Foundation. The subjects were highly selected in terms of both aptitude and motivation. Both sexes averaged above the ninety-seventh percentile in both the Mathematics and Verbal Preliminary Scholastic Aptitude Tests (using national norms for each sex). The average high school grade averages were over 93 for both sexes. Complete criterion and test data were obtained on 112 subjects (67 boys and 45 girls).

### *Tests and Measures*

The name, description, and hypothesized factor for each test are given below:

1. Advanced Vocabulary—Select the word

---

[1] The research here was completed in partial fulfillment of the degree of Master of Arts, Cornell University. Gratitude is expressed to Patricia C. Smith of Cornell University who directed the thesis, to Smith and Lorne M. Kendall of Cornell whose criticisms of this paper were most helpful, to the late T. R. Nielsen of Cornell who directed the Cornell National Science Foundation program and whose cooperation and help made this study possible, and finally to Matilda Kejner, Cornell University (Visiting Professor, University of Cordoba, Cordoba, Argentina), who was responsible for the scoring of the originality measures.

[2] In the previous paper (Locke, 1963) the classroom and out-of-class achievement factors were referred to as "structured" and "self-initiated" achievement, respectively.

that means the same from four alternatives (verbal; French, 1954).

2. Wide Range Vocabulary—Select the word that means the same from five alternatives (verbal; French, 1954).

3. Verbal Comprehension—Select the word that means the same from five alternatives (verbal; Guilford & Zimmerman, 1956).

4. Suffixes—Write as many words as possible ending with a certain suffix (word fluency; French, 1954).

5. Prefixes—Write as many words as possible beginning with a certain prefix (word fluency; French, 1954).

6. Topics—Think of as many ideas as possible about a given topic (ideational fluency; French, 1954).

7. Theme—Write a theme about a given topic (ideational fluency; French, 1954).

8. Thing Category—List as many things as possible that belong to a certain category (ideational fluency; French, 1954).

9. Consequences (obvious)—Write as many consequences as possible for a number of situations (ideational fluency; Christensen, Merrifield, & Guilford, 1958).

10. AC Test of Creative Ability (Form B, Part V)—Think of as many uses as possible for a number of objects. Quality was emphasized over quantity but only quantity was scored (ideational fluency; Harris & Simberg, undated).

11. AC Test of Creative Ability (Form A, Part II)—Think of as many reasons as possible why certain facts could be true (ideational fluency; Harris & Simberg, undated).

12. AC Test of Creative Ability (Form B, Part V)—Think of as many uses as possible for a number of objects (ideational fluency; Harris & Simberg, undated).

13. Categories test—Give as many similarities as possible between a number of objects or concepts (score was the total number of similarities given; creative ability; unpublished, author).

14. Associational Fluency I—Write words similar in meaning to other words (Christensen & Guilford, 1959).

15. Expressional Fluency—Write as many sentences as possible using words that begin with certain letters (Christensen & Guilford, 1959).

16. Pertinent Questions—Ask as many as four questions that should be considered in a number of situations (conceptual foresight; Berger, Merrifield, & Guilford, 1960).

17. False Premises—Indicate whether a number of syllogisms are "true" or "false" (deduction; French, 1954).

18. Word Squares—Fill in the missing word in a word matrix from a list of words, in order to fulfill a certain relationship between the words (deduction; French, 1954).[3]

19. Letter Grouping—Indicate which group of four letters does not follow the same rule as the other three (induction; French, 1954).

20. Marks—Put a mark in the proper space in the sixth row of a matrix according to the rule used in the first five rows (induction; French, 1954).

21. Ship Destination—Calculate the distance between two points ("ship" and "port") in a matrix in a number of situations given certain rules about how to take into account wind, current, and heading (general reasoning; Christensen & Guilford, 1956).

22. Special Problems—Solve a number of problems thought to require ingenuity (creative reasoning; Smith & Smith, 1956).

23. General Reasoning (Guilford & Zimmerman, 1956).

24. Bennett Mechanical Comprehension Test (Form CC).

25. Minnesota Clerical Test (speed; see text).

26. SRT Scale—Paired comparison adjective check list (scientific research temperament; Kosinar, 1960).

27–30. Gordon Personal Profile—Yields measures of Ascendancy, Responsibility, Emotional Stability, and Sociability.

31–34. Gordon Personal Inventory—Yields measures of Cautiousness, Original Thinking, Personal Relations, and Vigor.

The Minnesota Clerical Test was given with no time limit, and the amount of time spent on it was taken as a measure of speed. The total scores from the Gordon Profile and Inventory were eliminated to avoid spurious correlations between the totals and the subscores.

For the AC Test of Creative Ability (Form B, Part V) quality instructions were given to see whether this would affect the correlations between this and the other ideational fluency measures. Both forms of Part V of the AC test were also scored for Originality. These scores were obtained by weighting each response according to its frequency in the group, the more infrequent responses being given more weight. These weights were summed for all responses and divided by the number of categories used by the subject to attempt to control for fluency. These average category originality scores will be referred to as Originality A and B.

---

[3] A third deduction test from the French kit was eliminated due to extreme restriction of range. Over 50% of the subjects obtained perfect scores on this test.

The following 11 additional measures were also obtained for each subject from questionnaires:

35. Number of hours per week spent on independent study at high school.

36. Number of unassigned books read per month at high school.

37. Socioeconomic status based on a crude rating of the father's job level on a 4-point scale.

38. Number of science courses offered by each high school.

39. Size of home city.

40. The size of high school senior class.

41–45. Opportunities for research based on five questions asking the subject how much time and opportunity he had for independent work and research at high school.

## RESULTS

The above 45 variables were intercorrelated using all 122 subjects. The intercorrelation matrices were not computed separately for both sexes at this point because the sample size was so small compared to the number of variables that there would be considerable sampling and random error in such matrices. However, in an earlier analysis of these data (Locke, 1962), using a slightly different procedure, the factor structures were found to be quite similar for the two sexes. The resulting matrix was factor analyzed using the centroid method and rotated using the quartimax method on a Burroughs 220 computer. The Originality measures were not included in this analysis since they were found not to be completely independent of the fluency scores for the same tests. Twelve factors were extracted in the analysis. All loadings of .30 or greater (with two exceptions where loadings below .30 are shown) are shown below for each factor.

Factor I: Ideational fluency

| | |
|---|---|
| Consequences (obvious) | .66 [4] |
| AC (Form B, Part V) | .66 [4] |
| Categories | .64 [4] |
| Topics | .63 [4] |
| AC (Form A, Part II) | .62 [4] |
| AC (Form A, Part V) | .60 [4] |
| Thing category | .55 [4] |
| Pertinent questions | .54 [4] |
| Theme | .46 [4] |
| Associational fluency | .40 [4] |

All the open-ended tests where scores were calculated from frequency of response loaded highest on this factor with the exceptions of expressional fluency, suffixes, and prefixes which perhaps have more strictly to do with words than with ideas.

Factor II: Vocabulary

| | |
|---|---|
| Verbal comprehension | .86 [4] |
| Wide range vocabulary | .84 [4] |
| Advanced vocabulary | .84 [4] |
| Suffixes | .32 [4] |

This is clearly a vocabulary factor with a small loading for a word fluency test.

Factor III: General reasoning

| | |
|---|---|
| Special problems | .72 [4] |
| General Reasoning | .71 [4] |
| Ship Destination | .63 [4] |
| False premises | .60 [4] |
| Bennett (Form CC) | .57 [4] |
| Marks | .45 [4] |
| Original thinking | .43 |

The variables having high loadings on this factor all have to do with solving problems of a mathematical or logical nature. Both the General Reasoning and Ship Destination tests have been found previously to load on a general reasoning factor (Christensen & Guilford, 1956; Guilford & Zimmerman,

---

[4] Denotes that this variable loaded higher on this factor than any other. This notation will mean the same for all other factors.

1956). The reasoning and letter grouping tests surprisingly did not load on this factor.

Factor IV: Self-control versus impulsiveness

| | |
|---|---|
| Responsibility | .68[4] |
| Cautiousness | .68[4] |
| SRT | −.65[4] |
| Personal relations | .59[4] |
| Emotional stability | .50[4] |

This is a personality factor with high positive scores being associated with self-descriptions emphasizing responsible, cautious, controlled, and stable behavior, and low scores indicating impulsive, imaginative behavior and lack of self-discipline.

Factor V: School and city size

| | |
|---|---|
| Class size | .89[4] |
| City size | .84[4] |
| Number of science courses | .56[4] |
| Research opportunities—5 | .33 |

The three variables loading highest on this factor all have to do with the size and perhaps quality (i.e., number of courses offered) of the place where each subject went to high school.

Factor VI: Slowness

| | |
|---|---|
| Minnesota Clerical (time) | .65[4] |
| Letter grouping | −.58[4] |
| Marks | −.40 |
| Bennett (Form CC) | .33 |
| Associational Fluency | −.31 |
| Expressional Fluency | −.31[4] |

People scoring high on this factor spent a long time on the Minnesota Clerical Test and did poorly on the Inductive Reasoning tests.[5] The Bennett Mechanical Comprehension Test was given as a "take home" test with no time limit so the amount of time spent may have entered into the scores; this might explain its positive loading. The Associational Fluency and Expressional Fluency tests, both measures of "rate of word output", were timed and this could explain their negative loadings.

Factor VII: Research opportunities

| | |
|---|---|
| Research opportunities—1 | .68[4] |
| Research opportunities—5 | .56[4] |
| Research opportunities—2 | .54[4] |
| Research opportunities—3 | .30 |

This is clearly a research opportunities factor. Questions 1 and 2 concerned relations with teachers and their ability and willingness to help with independent research. Question 5 asked about the adequacy of library and laboratory facilities in the subject's school and home town.

Factor VIII: Orientation toward people

| | |
|---|---|
| Ascendancy | .81[4] |
| Sociability | .64[4] |

Both the ascendancy and sociability measures are concerned with liking for other people and groups, thus this factor was named "Orientation toward people."

Factor IX: Creative energy

| | |
|---|---|
| Original thinking | .61[4] |
| Vigor | .56[4] |

Since this factor suggests a combination of originality and motivation or energy, it was named "Creative energy."

Factor X: Independence versus socioeconomic status

| | |
|---|---|
| Socioeconomic status | −.39[4] |
| Hours on independent study | .32[4] |
| Ship Destination | −.29 |
| Number of unassigned books read | .27[4] |

---

[5] It should be mentioned that the induction tests were harder to finish for this group than were any of the other tests. This is one of the few tests on which no subject came even close to obtaining a perfect score or even finishing. This suggests that a considerable speed factor was present in the scores.

A high score on this factor seems indicative of independence, namely reading unassigned books and spending time doing unassigned high school work. Low scores indicate high socioeconomic status and general reasoning ability. This suggests that the subjects in this program who were from the lower class families (actually in this sample nearly all subjects' fathers were white collar so the range was between middle and upper-middle or upper classes), did more work on their own. This may reflect more drive on the part of these students as compared to those students with higher class backgrounds.

The factor is admittedly not a clear one, since two of the loadings are below .30.

### Factor XI: ?

| | |
|---|---|
| D-3 | −.40 |
| Research opportunities—3 | .40[4] |

This factor is not easily interpretable. The research opportunity question 3 concerned amount of parental encouragement for research.

### Factor XII: Speed fluency

| | |
|---|---|
| Research opportunities—4 | .44[4] |
| Thing Category | .39 |

This appears to be a speed factor. The research opportunity—4 question concerned the amount of time available for extra work; the Thing Category test, like the other fluency tests, involved giving as many ideas as possible about a topic in a given amount of time. Thus the factor seems to involve a combination of speed and high output of verbal material.

The three or four best measures of each factor were summed to form factor index scores after converting all raw scores to *z* scores. In all but one case (involving highly discrepant loadings of different variables in the same factor) all measures were weighted equally in forming index scores. The use of unit weights in the case where factor loadings do not vary too widely has some advantage over weighting by factor loadings, since the latter policy would bring up the problem of instability of the weightings in different situations due simply to differences in the type of factor analytic, and rotation procedures used, the number and type of tests used, and the number of factors extracted.

Table 1 shows the variables comprising each factor, the relative weight of each (where applicable), and the direction of each variable. Index scores were not computed for Factors 11 and 12 since they were neither easily interpretable nor internally consistent. In computing Index X (from Factor X) the Ship Destination test was not included since it was already in another factor. The "unassigned books read" measure was included, however, in spite of its low loading because it did seem to go logically with hours spent on independent work. Admittedly the factor is not very internally consistent. The Originality A and Originality B scores were summed to form an Originality index which will be numbered Index XIII.

One advantage of using index scores as opposed to individual test scores is that the former procedure reduces the ratio of variables to subjects and therefore the number of significant predictor-criterion correlations that could be obtained by chance. A second advantage, when the components of an index are highly intercorrelated, is that index scores will be more internally consistent and stable over time than single measures. Thirdly, index scores should contain less "specific" variance and more "common" variance than single tests. Finally, if the latter two conditions are

TABLE 1

DESCRIPTION OF FACTOR INDEXES

| Factor index | Name | Index components | Number variables |
|---|---|---|---|
| I | Ideational Fluency | Consequences + [AC-B-Part V] + Topics + [AC-A-Part II][a] | 4 |
| II | Vocabulary | Verbal Comprehension + Advanced Vocabulary + Wide Range Vocabulary | 3 |
| III | General Reasoning | Special Problems + General Reasoning + Ship Destination + False Premises | 4 |
| IV | Self-Control versus Impulsiveness | Responsibility + Cautiousness − SRT + Personal Relations + Emotional Stability | 5 |
| V | School and City Size | Class Size + City Size + ½ [Number of courses] | 3 |
| VI | Slowness | Minnesota Clerical (time) − Letter Grouping − Marks | 3 |
| VII | Research Opportunities | Questions 1 + 5 + 2 | 3 |
| VIII | Orientation towards People | Ascendancy + Sociability | 2 |
| IX | Creative Energy | Original Thinking + Vigor | |
| X | Independence versus Socioeconomic Status | − [Ses] + [Hours on independent work] + [Number of unassigned books read] | 3 |
| XIII | Originality [b] | Originality A + Originality B | 2 |
| | Total Number Variables | | 34 |

[a] The Categories test should have been in this index since it had the third highest loading on Factor I, but it was omitted due to a clerical error. Since all the factor loadings on this factor were very similar, the difference is undoubtedly trivial.

[b] Not in original factor analysis; see text for explanation.

satisfied, index scores should give more consistent results with different samples, although in a given situation an individual test might predict better than an index score.

The reliabilities (internal consistencies) of the indexes described in Table 1 are given in Table 2.[6]

The 11 factor indexes and the 2 criterion factor scores[7] were intercorrelated

TABLE 2

INTERNAL CONSISTENCIES OF FACTOR INDEXES

| Factor index | Average intercorrelation [a] of component tests | Internal consistency of index score |
|---|---|---|
| Ideational Fluency | .40 | .73 |
| Vocabulary | .74 | .90 |
| General Reasoning | .43 | .82 |
| Self-Control versus Impulsiveness | .36 | .85 |
| School and City Size | .58 | .81 |
| Slowness | .37 | .64 |
| Research Opportunities | .34 | .61 |
| Orientation toward People | .54 | .70 |
| Creative Energy | .43 | .60 |
| Independence versus Socioeconomic Status | .19 | .41 |
| Originality | .30 | .46 |

[a] Averaging was done after appropriate $z$ transformations.

---

[6] The internal consistencies were computed by means of the familiar Spearman-Brown formula

$$r_{aa} = \frac{a\bar{r}}{1 + (a - 1)\bar{r}}$$

where $a$ is the number of tests or forms of the test and $\bar{r}$ is the average intercorrelation among the tests.

[7] Factor scores for the criterion factors were computed by weighting each variable according to its factor loading (see Locke, 1963). This was necessitated by the wide variation in factor loadings of the variables in each factor.

TABLE 3

INTERCORRELATIONS OF INDEX SCORES

| Index | I | II | III | IV | V | VI | VII | VIII | IX | X | XIII |
|---|---|---|---|---|---|---|---|---|---|---|---|
| I Ideational Fluency | — | .07 | .05 | −.06 | −.03 | −.25 | −.03 | .19 | .24 | .14 | .40 |
| II Vocabulary | | — | .17 | −.03 | .12 | −.14 | .19 | .00 | .04 | .16 | .07 |
| III General Reasoning | | | — | −.09 | .07 | −.30 | −.04 | .10 | .29 | −.06 | .03 |
| IV Self-Control versus Impulsiveness | | | | — | −.10 | −.04 | .04 | −.13 | .18 | −.24 | −.09 |
| V School and City Size | | | | | — | −.08 | .15 | .06 | −.06 | .08 | .03 |
| VI Slowness | | | | | | — | −.08 | −.04 | −.22 | −.02 | −.10 |
| VII Research Opportunities | | | | | | | — | .06 | .12 | −.19 | .06 |
| VIII Orientation toward People | | | | | | | | — | .26 | −.05 | .11 |
| IX Creative Energy | | | | | | | | | — | .07 | .15 |
| X Independence versus Socioeconomic Status | | | | | | | | | | — | .18 |
| XIII Originality | | | | | | | | | | | — |

TABLE 4

CORRELATIONS OF INDEX SCORES AND CRITERION FACTORS
($N = 112$)

| Factor index | Criterion factor | | | | | |
|---|---|---|---|---|---|---|
| | Classroom achievement | | | Out-of-class achievement | | |
| | Total | Boys | Girls | Total | Boys | Girls |
| Ideational Fluency | .03 | .09 | −.07 | .05 | .02 | .11 |
| Vocabulary | .34** | .54** | .08 | .13 | .01 | .32* |
| General Reasoning | .18 | .27* | −.13 | .06 | .01 | −.12 |
| Self-Control versus Impulsiveness | .22* | .28* | .07 | −.09 | .00 | −.29* |
| School and City Size | .06 | .07 | −.06 | .32** | .30* | .29$^a$ |
| Slowness | −.06 | −.16 | .07 | .06 | .02 | .05 |
| Research Opportunities | .18 | .21 | .09 | .17 | −.01 | .35* |
| Orientation Toward People | −.01 | −.06 | .05 | .02 | .04 | −.05 |
| Creative Energy | .14 | .38** | −.35* | .25** | .23 | .22 |
| Independence versus Socioeconomic Status | −.21* | −.14 | −.38** | .25** | .16 | .44** |
| Originality | .02 | .08 | −.20 | .31** | .26* | .38** |
| R | .50$^b$ | .70$^c$ | .46$^c$ | .58$^b$ | .37$^c$ | .66$^c$ |
| $R_c$ | .44* | .68** | .41* | .50** | .34* | .59** |

[a] This correlation is the same as the one directly above when rounded but does not actually reach significance at the .05 level.

[b] Based on all factor indexes.

[c] Based on indexes having significant correlations ($p < .05$ or better) with criterion in question for group in question, including school and city size for girls for the out-of-class achievement factor.

* $p < .05$.

** $p < .01$.

for: the total group, all boys, and all girls. Multiple *R*s were computed for: all 11 index scores against both criterion scores for the total group, all index scores which correlated significantly with each criterion for the boys, and all index scores which correlated significantly with each criterion for the girls. All multiple *R*s were corrected for bias using Guilford's (1956, p. 399) correction formula.

Table 3 shows the intercorrelations of the index scores for the total group. The "validity coefficients" for the total group, boys, and girls, and the multiple *R*s described above are shown in Table 4.

For the total group the variables significantly related to classroom achievement were vocabulary, self-control, and socioeconomic status. It can be seen from an examination of the correlations by sex that the total group correlations were due mainly to the boys. The only measures significantly related to classroom achievement for the girls were lack of creative energy and socioeconomic status.

Creative energy was positively related to classroom achievement for the boys, the opposite direction from the relationship obtained with the girls. No obvious explanation of this finding can be given but a possible interpretation will be suggested. In the first place, the original thinking measure, which is a component of the creative energy index, would be more aptly called "intellectuality" since most of the items deal with curiosity and interest in knowledge and critical or philosophical thinking rather than originality per se. It is possible that males who are "vigorous intellectuals" are encouraged to achieve in the classroom by existing male social norms emphasizing scholastic "achievement," whereas females who are vigorous intellectuals are actively discouraged from intellectual pursuits by existing female social norms emphasizing "popularity." The energy of the female intellectuals may be channeled off into other (extracurricular) activities thus detracting from their required school work. The above explanation is, of course, highly speculative and cannot actually be tested with the data from this study.

The probable reason why the girls were less predictable on this factor than the boys was that there was severe restriction of range in the girls' groups on classroom achievement as compared to the boys' group. The variances of the two groups on this factor are significantly different, ($F = 1.85$, $p < .05$), although the means of the two groups are nearly identical.

There is somewhat more similarity between the boys and girls on the out-of-class achievement factor. For both sexes and the total group originality was significantly related to this criterion. For the total group and the boys, school and city size was significantly related to out-of-class achievement, and the correlation for the girls' group just misses reaching significance at the .05 level. In addition, for the girls' group, vocabulary, impulsiveness, research opportunities, and independence were significantly related to out-of-class achievement. The most unexpected result for the girls was the correlation of vocabulary with this criterion factor. There is no obvious explanation for this finding (plotting of the relationships yielded no curvilinear trends), though it should be noted that the correlations of vocabulary with the two criterion factors were not significantly different for the women. This was decidedly not the case for the men, where the correlation of vocabulary with classroom achievement was significantly higher ($p < .05$) than the correlation with out-of-class achievement. It might also be noted that the partial correlation of vocabulary with out-of-class achievement for the girls is only .24 while the partial correlation with classroom achievement is .23. Thus when all other factor indexes are controlled, there is no significant relationship between vocabulary and either criterion factor for the girls.

The greater predictability of the girls than the boys on the out-of-class achievement factor cannot be explained on the basis of differences in the means or variances of the two groups on the criterion factors since the groups were nearly identical in these respects.

The significant relationship between research opportunities and out-of-class achievement for the girls has a possible explanation. It will be recalled that two of the three questions involved in this index concerned teachers' willingness to help the student with independent research. It is commonly accepted that personal relationships are more important for women than for men and thus the influence of teachers may have been more important in motivating girls to do extra research than for boys.

It is probable that the obtained relationship between school and city size and out-of-class achievement was due to the fact that the larger high schools and cities have better facilities (libraries, laboratories, etc.), better teachers, more frequent science fairs, and a generally more "intellectual" orientation than smaller high schools and towns.

All the corrected multiple *R*s in Table 4 are significant at the .05 level or better. It should be noted, however, that these "predictors" are not all "independent variables." All the out-of-class achievement behavior took place *before* any of the students were tested and in addition the number of hours spent on independent work and the number of unassigned books read were behaviors that were undoubtedly occurring *along with* the out-of-class achievement behavior. Similarly, part of the classroom achievement behavior (high school grades and summary ratings) took place before the students were tested and part of it (Cornell grades and summary ratings) took place after testing. The multiple *R*s merely indicate the amount of criterion variance "accounted for" in variables not included in the criterion measures. All the criterion measures and all predictor indexes could however be used to try to predict later (college and postcollege) achievements of these subjects.

## DISCUSSION

Since the two criterion factors used in this study were independent, it was to be expected that different measures would be related to each. Insofar as the classroom achievement factor is similar to what is traditionally called "grade achievement" and the out-of-class achievement factor is similar to what other researchers have referred to as "creativity," this study can be compared to other studies which have explored the correlates of these factors.

Holland (1961), in a study of National Merit Scholars, found lack of independence, self-control, and math scholastic aptitude to be significantly related to science grades. Smith (1959) found significant correlations between grades and two measures from the Gordon Personal Profile which comprised part of the self-control index used in this study. Tallent (1956) also found a measure of self-control to be related to grades. Numerous studies, of course, have found relationships between mathematical and verbal IQ measures and grade achievement. The findings of this study would seem to agree with previous findings in that math and verbal IQ (vocabulary and general reasoning) and self-control were significantly related to classroom achievement, at least for males. The relationship found for the girls between lack of independence (high socioeconomic status) and classroom achievement also parallels Holland's finding above.

It is not surprising that achievement in the structured environment of the classroom should be related to intelligence test scores since the types of problems encountered in the two situations are often similar. Similarly the relation between self-control or responsibility and classroom achievement is not surprising since the latter often requires a certain amount of conformity to standards, the following of directions, and faithful compliance to norms set by teachers. The point is not to consider unimportant this form of achievement simply because it requires "conformity," but rather to appreciate the fact that classroom achievement is only one kind of achievement requiring a particular pattern of traits and abilities that may be inappropriate or irrelevant to other forms of achievement.

Holland (1961), in the study described above, found lack of responsibility, independence, lack of desire for status, originality, and initiative to be significantly related to a creativity criterion much like the out-of-class achievement factor used here. Piers, Daniels, and Quackenbush (1960) found an originality measure to be related to creativity ratings. These findings also parallel the ones in this study. Impulsiveness (or lack of self-control and responsibility), independence (or low socioeconomic status) and originality correlated with out-of-class achievement for the females and originality for the males. It should also be noted that two measures of initiative formed part of the out-of-class achievement factor in this study (Locke, 1963).

Again it is not surprising that achievement outside of the structured environment of the classroom should be more highly related to personality measures indicative of independence and initiative and intellectual factors such as originality than to traditional intelligence measures and self-control. Outside the classroom a student is told neither what to do nor how to do it; the discovery of both means and ends must usually be made independently or with the help of a teacher contacted at the student's own initiative. Again it should not be assumed that there is anything inherently superior in out-of-class achievement or creative achievement as Holland (1961) calls it. Little is known about the relationship between the two criterion factors used here and later life success, however defined. Similarly, little is known about the kinds of measures most highly predictive of various kinds of later life success. It may be that schools will find it more profitable in the long run to nurture abilities other than intelligence and utilize criteria of success other than grades.

It is perhaps fortunate that the two measures of achievement used in this study (and the similar indexes used by Holland) appear to be independent rather than negatively related. This would suggest that there is nothing to prevent schools and teachers from emphasizing *both* criteria rather than emphasizing one at the expense of the other.

## REFERENCES

BERGER, R. M., MERRIFIELD, P. R., & GUILFORD, J. P. *Pertinent questions.* Beverly Hills, California: Sheridan Supply, 1960.

CHRISTENSEN, P. R., & GUILFORD, J. P. *Ship Destination Test.* Beverly Hills, California: Sheridan Supply, 1956.

CHRISTENSEN, P. R., & GUILFORD, J. P. *Christensen-Guilford Fluency Tests.* Beverly Hills, California: Sheridan Supply, 1959.

CHRISTENSEN, P. R., MERRIFIELD, P. R., & Guilford, J. P. *Consequences.* Beverly Hills, California: Sheridan Supply, 1958.

FRENCH, J. W. (Ed.) *Kit of selected tests for reference aptitude and achievement factors.* Princeton: Educational Testing Service, 1954.

GUILFORD, J. P. *Fundamental statistics in psychology and education.* New York: McGraw Hill, 1956.

GUILFORD, J. P., & ZIMMERMAN, W. S. *Guilford-Zimmerman aptitude survey.* Beverly Hills, California: Sheridan Supply, 1956.

HARRIS, R. H., & SIMBERG, A. *AC Test of Creative Ability.* Flint, Michigan: AC Spark Plug Division, General Motors Corporation, undated.

HOLLAND, J. L. Creative and academic performance among talented adolescents. *J. educ. Psychol.,* 1961, **52,** 136–147.

KOSINAR, W. C. *The SRT Scale.* Chicago: Psychometric Affiliates, 1960.

LOCKE, E. A. Patterns of achievement of gifted science students. Unpublished master's thesis, Cornell University, 1962.

LOCKE, E. A. The development of criteria of student achievement. *Educ. psychol. Measmt.,* 1963, **23,** 299–307.

PIERS, E. V., DANIELS, J. M., & QUACKENBUSH, J. F. The identification of creativity in adolescents. *J. educ. Psychol.,* 1960, **51,** 346–351.

SMITH, D. D. Traits and college achievement. *Canad. J. Psychol.,* 1959, **13,** 93–101.

SMITH, O. W., & SMITH, PATRICIA C. Validation of selection devices for designers. Unpublished study, Cornell University, 1953.

TALLENT, N. Behavioral control and intellectual achievement of secondary school boys. *J. educ. Psychol.,* 1956, **47,** 490–503.

(Received January 14, 1963)

1963, Vol. 54, No. 5, 249–252

# EFFECT OF INTRALIST ITEM SIMILARITY ON PAIRED-ASSOCIATE LEARNING BY LEARNERS OF HIGH AND LOW VERBAL ABILITY[1]

LOWELL SCHOER
*State University of Iowa*

The present study was designed to determine whether differential susceptibility to the inhibition introduced by competing responses may account for at least part of the difference in the performance of high and low verbal ability learners. The learning material consisted of 2 lists of paired adjectives: List I with low intralist response competition and List II with high intralist response competition. A significant interaction between ability level and list was found. This finding suggests that the low verbal ability learner is more adversely affected than the high verbal ability learner by response competition.

The influence of individual differences in verbal ability on learning performance has been widely investigated, but few attempts have been made to determine those factors which might account for this influence. The present study was designed to investigate one such factor.

One variable that influences paired-associate learning is the similarity among the stimulus words (Underwood, 1951, 1953) and, although perhaps to a lesser degree, the similarity among the response words (Feldman & Underwood, 1957). Since discrimination among the stimulus words and among the response words seems a necessary part of paired-associate learning (Gibson, 1940), any increase in difficulty as a function of increased similarity among stimuli and among response words is quite probably at least partially due to the more difficult discriminations these similar words require the subject to make.

As a result of this increased difficulty in discrimination there is likely to be increased response competition among the words in the list due to stimulus and response generalization. The question investigated in the present study was whether subjects of high and low verbal ability are affected differently by such an increase in response competition. If such an effect were found, it would suggest that perhaps one of the factors which accounts for the difference in the learning performance of high and low verbal ability subjects is differential susceptibility to the inhibition introduced by response competition.

Investigating this possibility by comparing the performance of subjects of high and low verbal ability on lists with varying degrees of similarity among stimulus and response words involves the assumption that an increase in such similarity increases response competition. A necessary part of the present study then was to determine the validity of this assumption. This was done in a preliminary analysis of the data, using number of intralist intrusions as a measure of response competition.

## PROCEDURE

The subjects for the experiment came from the undergraduate course in educational psychology at the State University of Iowa. Early in the semester all the students in one section of the course were given a 75-item vocabulary test. Those who scored in the upper half of the class on this test constituted the pool from which the high ability experimental subjects were drawn, and those who scored in the lower half constituted the pool from which the low ability experimental subjects were drawn.

---

[1] The author wishes to thank W. Lucht and G. Nicholsen for their aid in gathering the data for this experiment and J. B. Stroud for his helpful suggestions in preparing the manuscript.

The learning material consisted of two lists of paired adjectives taken from the Haagen (1949) list. List I was composed of eight pairs of adjectives of high familiarity with low association between the stimulus and response members of each pair as well as among the eight stimulus members and eight response members of the list. List II consisted of eight pairs of adjectives matched on the dimension of familiarity with the words in List I. Also like the words in List I there was low association between the stimulus and response members of each pair. Unlike List I, however, all eight of the stimulus words were synonymous and all eight of the response words were synonymous. It was anticipated that the synonymity among the stimulus and among the response words in List II would provide considerably more intralist response competition than would be present in List I where the eight stimulus words were unrelated to each other as were the eight response words.

The learning sessions for the experimental subjects began approximately 3 weeks after the vocabulary test was given. The learning material was presented to the subjects in individual sessions. A Hull-type memory drum was employed using a 2-second exposure of the stimulus and a 2-second exposure of the stimulus and response together. The subject was asked to anticipate the response before it appeared. There was a 4-second interval between trials. All subjects were given 3 trials on a practice list of three pairs of adjectives prior to learning the experimental list. Each subject was given 12 learning trials on the experimental list and learned only one such list.

## RESULTS

Table 1 shows a decided increase in intrusions as a function of increased similarity among the stimulus and among the response words. The primary interest in Table 1 is in the main effect of list, and analysis of variance showed this effect to be significant at well beyond the .01 level. This significant main effect lends considerable support to the assumption that an increase in stimulus and in response similarity increases response competition.

TABLE 1

NUMBER OF INTRUSIONS ON LISTS I AND II BY ABILITY LEVELS

$(N = 52)$

| Number of intrusions | Ability level High $(n = 26)$ | Ability level Low $(n = 26)$ |
|---|---|---|
| List I | 14.2 | 13.5 |
| List II | 24.8 | 31.7 |

As can be seen by looking at the mean raw scores given in the parentheses in Table 2, List II was considerably more difficult than List I. In order to put the scores of Lists I and II on the same scale, the raw scores were transformed to standard scores and the analysis was based on these standard scores. This transformation was made by combining scores for all the subjects on List I and determining $z$ score equivalents for each raw score value. In order to remove the decimals and the negative signs from these $z$ scores, each was multiplied by 10 and added to 50. The same procedure was followed with List II.

These standard scores were then analyzed using a Treatments × Levels statistical design (Lindquist, 1953). The main effect for ability was significant at greater than the .01 level. Because of the nature of the transformation employed the main effect for "list" would not be expected to be significant.

The primary interest for the purposes of the present study was in the interaction between ability and list; this interaction is significant at greater than the .05 level. The source of this interaction is evident from Table 2. A test of the simple effects in Table 2 indicates that the difference between the high and low ability students is not significant on

TABLE 2

MEAN STANDARD SCORES OF HIGH AND LOW ABILITY SUBJECTS ON LISTS I AND II

($N = 52$)

| List | Ability level |  |
|---|---|---|
|  | High ($n = 26$) | Low ($n = 26$) |
| I |  |  |
| Mean standard scores | 51.5 | 49.2 |
| Mean total correct responses | 61.0 | 56.5 |
| II |  |  |
| Mean standard scores | 55.3 | 44.8 |
| Mean total correct responses | 38.7 | 21.5 |

List I ($t = .84$, $p > .10$) but is highly significant on List II ($t = 3.47$, $p < .01$).

## DISCUSSION

The present study suggests that one of the factors that may account for at least part of the difference in the performance of high and low ability learners is differential susceptibility to the inhibition introduced by competing responses. This indicates the possibility that high and low ability students might perform equally well on a learning task if the material to be learned is presented in such a way as to minimize the possibility that strong incorrect and competing responses are aroused in the learner.

This may, in fact, be the reason that high and low ability learners do equally well when learning material is "programed," since in linear programing a conscious effort is made to present learning material in such a way as to make the correct response relatively obvious, i.e., minimize the possibility that plausible incorrect responses will be aroused in the learner. To the extent that this reduces response competition it reduces one of the factors that seems to cause at least part of the difference in the performance of high and low ability learners and should, therefore, make their learning performance more similar.

This possibility finds some support in the results of a recent study by Shay (1961). Varying step size, Shay found a significant interaction between ability and step size, with the above average learners showing significantly less decrease in probability of correct response as step size is increased than did the below average learner. The added response competition introduced as the steps in the program becomes larger could account for at least part of this effect.

The suggestion that the less probability there is of a stimulus arousing two or more competing responses in the learner the less probability there is that there will be a difference in the performance learners of high and low verbal ability may also indicate the reason why such a low relationship is typically found between intelligence and performance in paired-associate learning. The learning material employed in such experiments has typically been of the type used in List I, i.e., intralist response competition is at a minimum. To the degree that this type of material reduces response competition, which at least partially accounts for performance differences between high and low ability learners, it would tend to reduce the differences found between the performance of high and low ability learners.

## REFERENCES

FELDMAN, S. M., & UNDERWOOD, B. J. Stimulus recall following paired-associate verbal learning. *J. exp. Psychol.*, 1957, **53**, 11–15.

GIBSON, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

HAAGEN, C. H. Synonymity, vividness, familiarity and association value of 400 pairs of common adjectives. *J. Psychol.*, 1949, **27**, 453–463.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

SHAY, C. B. Relationship of intelligence to step size on a teaching machine program. *J. educ. Psychol.,* 1961, **52**, 98–103.

UNDERWOOD, B. J. Studies of distributed practice: II. Learning and retention of paired-adjective lists with two levels of intra-list similarity. *J. exp. Psychol.,* 1951, **42**, 153–161.

UNDERWOOD, B. J. Studies of distributed practice: IX. Learning and retention of paired adjectives as a function of intra-list similarity. *J. exp. Psychol.,* 1953, **45**, 143–149.

(Received March 13, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 5, 253–260

# COMPARISON OF SEVERAL RESPONSE MODES IN A REVIEW PROGRAM

JOANNA P. WILLIAMS
*University of Pennsylvania*

A program was administered as review to college students in 4 versions, which varied in response mode. The performance of the constructed-response group and the multiple-choice group did not differ, and these groups scored higher than the emphasis group, which read through the completed program with key words underlined, and the reading group, which read it without underlines. When the test item required that S respond with novel, "technical" terminology, the constructed-response group scored higher than the multiple-choice group. In terms of learning efficiency the groups were ordered as follows: emphasis (most efficient), reading, multiple-choice, constructed-response. Performance also varied with aptitude.

One of the most generally accepted principles in programed learning is that the student must actively participate in the instruction. This recommendation reflects the general psychological principle that a response must be made before it can be reinforced, and it is consistent with the findings of the older (preprograming) literature on active participation (e.g., Gates, 1917).

Despite the reasonable expectation that so important and general a principle would be easily demonstrated in the programed instruction context, the empirical support is surprisingly slim. Holland (1960) found that students who wrote an answer to each item made only half as many errors on a subsequent test as did students who read through already completed items. Krumboltz and Weisman (1962) found that subjects who wrote out responses scored higher than those who made covert responses and those who simply read the program, but these differences appeared only on a delayed criterion test and not on an immediate one. Yet the majority of studies (e.g., Evans, Glaser, & Homme, 1960; Keislar & McNeil, 1962) have not supported this principle.

Indeed, it is still an open question as to what constitutes active participation. Skinner (1958) suggests that the student be required to construct his own responses and write them in on each frame. Pressey (1950), on the other hand, suggests that simply to select a response from a set of alternatives would satisfy the requirements of active participation. Comparisons of the effectiveness of these two modes of presentation have usually failed to show differences (e.g., Coulson & Silberman, 1960; Hough, 1962). Fry (1960), however, using a Spanish vocabulary program found that a constructed-response mode led to higher achievement than did a multiple-choice mode on a recall test, though not on a test of recognition. There is no case where a typical continuous discourse program showed such an effect.

These two modes are not the only possibilities for ensuring active participation. Perhaps a subject need not construct each response nor even select it from an array, but simply focus his attention on it. In other words, the key points might be written into the program, but emphasized (by underlining, for example) in such a way that the subject's attention was ensured. This emphasis might lead to minimal but sufficient active participation.

Each of the presentation modes described might be considered as differing in the number of aspects of active participation which they include: the constructed-response mode contains choice and attention along with construction of the responses, the multiple-choice mode contains attention and choice, and the emphasis group includes only attention. From a theoretical standpoint it seems important to investigate the effectiveness of active participation and the particular aspects of it which are critical, and there are practical considerations to the problem as well. The cost of a program, the time it requires, the problems of administration, and the complexities of the teaching machine, if one is used, would all be cut down if a relatively simple mode of presentation were acceptable.

The purpose of this experiment was to compare the effectiveness of programs which differ in the extent of participation required of a subject. Four versions of the Holland-Skinner program in elementary psychology were developed: a constructed-response version, a multiple-choice version, an already completed version in which important portions of the frame were underlined for emphasis, and an already completed version which required no active participation in any form.

In addition to an assessment of effectiveness, as reflected in mastery of the material, the efficiency of the several versions was examined. While a particular presentation mode may result in a higher level of mastery, it may cost a great deal in extra time, and in certain situations might be rejected on the grounds of inefficiency. A characteristic finding (e.g., Goldbeck & Campbell, 1962) has been that training is more efficient when constructed responses are not required. In the present study, time scores were recorded in order to examine the relative efficiency of the particular presentation modes used.

There is a possibility that differences between the constructed-response mode and the multiple-choice mode are often obscured by the fact that the constructed-response mode is superior only for certain kinds of items. For example, if the response required on the test is a novel term which has been introduced by the program, or any word which is not a highly overlearned response per se, then presumably the overt practice from writing it out should aid learning. On the other hand, where the words required are common terms in the subject's vocabulary, constructing them within the context of the program may not be of much value. In other words, in the first case, when there is a novel term, there is response learning to be accomplished as well as the association of the response and the stimulus (the item), whereas in the second case, the response learning has been completed outside the context of the program and need not be provided for. The present experiment will test the hypothesis that the difference between the constructed-response mode and the multiple-choice mode depends on the type of test item.

In this experiment active participation was investigated in the context of review materials. To answer the question of whether a program written as introductory material could be useful as review material, a number of subjects were administered the criterion tests without having gone through the program in any form. While programed instruction appears to be a promising solution to the problem of equating the subjects on previous learning without taking up class time, an examination of its value as review is desirable before it is adopted wholesale by the schools.

While the major focus of the experiment was on active participation, it was possible to analyze the data with regard to another issue as well. The literature is unclear as to how the effectiveness of programed instruction is affected by aptitude. While typically aptitude is correlated with achievement in "conventional" instruction, Ferster and Sapon (1958) and Porter (1959) found no such relationship between aptitude and achievement when programed materials were used. In Porter's study, the subjects who received more typical instruction did show a correlation between aptitude and achievement. Lambert, Miller, and Wiley (1962), on the other hand, over a wide IQ range, found that with programed materials achievement scores were higher for high IQ groups. The present study will also examine the effectiveness of programed instruction as a function of aptitude, as measured by the Scholastic Aptitude Test-Verbal (SAT-V).

## METHOD

### *Subjects*

The subjects were 128 educational psychology students, all of whom had completed a semester of general psychology. A total of 35 additional subjects was dropped from the sample for the following reasons: failure to complete the training and testing during the scheduled class period, failure to follow directions, and unavailability of aptitude test scores. Eight subjects were dropped (randomly) in order to equate *N* in each group. The subjects were run in two replications: 56 out of 75 were included in the first replication, and 72 out of 96 in the second.

### *Materials*

Sets 7–11 (totaling 192 frames) of the Holland-Skinner program, The Analysis of Behavior, were reproduced in booklet form. Each frame appeared on a separate page, with the correct response (where required) printed alone on the following page. Each set of frames was stapled into a separate booklet. The program was reproduced in four forms:

1. Constructed response (CR): Standard format, in which the subject was required to write out answers on every frame. The subject found the correct response to each frame on the following page. Turning the page again exposed the next frame.
2. Multiple choice (MC): Under each item, presented in the standard way, two responses were printed. The subject was instructed to underline his choice, and he found the correct response on the following page.
3. Emphasis (E): Each item appeared as a complete statement; the words which in the standard program are left to the subject to fill in were underlined. No answer pages were provided.
4. Reading (R): Each item appeared as a complete statement, no part of which was underlined. No answer pages were provided.

An objective test was made up of 20 items taken from Test Unit 17 of the program. A total of 29 constructed responses was required of the subject. In addition, an "essay" test was given. The subjects were asked to write answers to the following two questions:

1. One-year-old Bobby whines a great deal, especially when he is hungry. What are two ways in which you might go about eliminating his whining? Describe them fully, using technical terms whenever possible.
2. Why is money so effective in producing desirable behavior? Analyze its function, using Skinnerian principles and terminology.

### *Procedure*

The subjects were randomly assigned to one of the four training conditions. On the first day of training, Sets 7, 8, and 9 were completed, and the following day, Sets 10 and 11 were completed. Each subject recorded the amount of time he had taken to work through the program. The subjects were excused from class when they had completed the assignment.

The criterion tests were administered 1 week after the start of training. In both replications the essay test was given prior to the objective test. Twenty minutes were allowed for each test, and all subjects finished within the time limits.

Before the experimental training the subjects were told that this was an experiment in programed instruction and that the class would discuss the data later in the semester. It was pointed out that their test scores would not count as part of their course grade, but that later lectures would assume mastery of the material presented in the program.

Another group ($N = 32$) of similar subjects who had also studied general psychology was given the criterion tests without having gone through the program in any form. Testing conditions were kept as similar as possible to those of the groups in the main design.

## RESULTS

Because there were no differences between the two replications in the mean number of correct responses on the objective test as a function of response mode ($F = 3.73$, $df = 1/120$) nor in aptitude ($F = 2.92$, $df = 1/120$), the two replications were combined for analysis.

Means and standard deviations of the scores are presented in Table 1. The standard deviations increase as the amount of participation decreases, and that of Group CR is significantly smaller than that of any of the other three groups (e.g., CR versus MC: $F = 2.51$, $df = 31/31$, $p < .02$; there were no differences among groups MC, E, and R). To offset the heterogeneity of variance, a significance level of .025 was set on subsequent statistical tests.

Table 2 indicates the frequency of scores within various portions of the range. It is evident that the smaller variability of Groups CR and MC is due to the absence of low scores in these groups.

The subjects in each of the four conditions were dichotomized at the median on the basis of aptitude score (SAT-V), and analysis of variance showed that there were no differences in aptitude among the four response modes. Figure 1 presents the mean score on the objective test as a function of response mode and aptitude level. Both main effects were significant (response mode: $F = 6.48$, $df = 3/120$, $p < .001$; aptitude: $F = 8.26$, $df = 1/120$, $p < .005$), and the interaction was not significant. Specific comparisons indicated that Groups CR and MC were not different from each other, but they scored significantly higher than did Groups E and R, ($F = 38.20$, $df = 1/120$, $p < .001$), which were also not significantly different from each other.

TABLE 1

NUMBER CORRECT ON THE CRITERION TEST

| Group | *M* | *SD* |
|---|---|---|
| CR | 23.5 | 2.1 |
| MC | 23.0 | 3.3 |
| E | 20.7 | 3.9 |
| R | 20.6 | 4.3 |

TABLE 2

FREQUENCY OF SCORES FALLING IN DIFFERENT PORTIONS OF THE RANGE

| Response mode | Score | | |
|---|---|---|---|
| | 10–15 | 16–21 | 22–27 |
| CR | 0 | 6 | 26 |
| MC | 1 | 6 | 25 |
| E | 3 | 14 | 15 |
| R | 5 | 11 | 16 |

The mean of the group which did not go through the program in any form was 8.8; the standard deviation was 5.0. The difference between this group and the other four is so large that a statistical test would be only a formality.

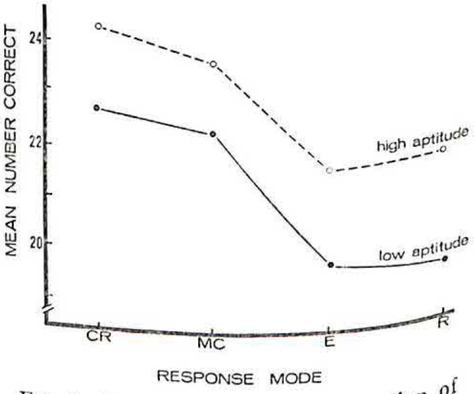


FIG. 1. Test performance as a function of response mode and aptitude.

The essay tests were graded on an 8-point scale (with 1 representing the best tests). Since the reliability of the two judges who scored these tests (advanced graduate students in psychology) was only .55, no analysis of these data has been made.

Figure 2 indicates the mean time taken to complete the training as a function of presentation mode and aptitude. Mode of presentation was a significant effect ($F = 65.15$, $df = 3/120$, $p < .001$), and each of the four groups was significantly different from the others. High aptitude subjects took significantly less time than low aptitude subjects ($F = 8.06$, $df = 1/120$, $p < .01$). There was no interaction between presentation mode and aptitude.

Correlations among aptitude, time taken to complete the training, and score on the objective test are presented in Table 3. In each of the four conditions, there was a positive correlation between aptitude and test score. Since there were no differences among these correlations, an overall estimate of the correlation was made; it was .38, significant at the .01 level. The correlation between aptitude and score in the control group (those subjects who did not receive any training) was .27, not significantly different from the other groups. In all conditions the correlation between aptitude and training time was negative, and in Group R this correlation reached significance. There was no relation between training time and test score. Each of these correlations was also computed with the third variable partialed out, and this procedure did not change the general pattern of correlations nor the conclusions based upon them.

TABLE 3

CORRELATIONS AMONG APTITUDE, TRAINING TIME, AND SCORE ON CRITERION TEST

| Group | Aptitude and score | Time and score | Aptitude and time |
|---|---|---|---|
| CR | .57* | −.17 | −.23 |
| MC | .23 | .31 | −.21 |
| E | .34* | −.14 | −.20 |
| R | .37* | .14 | −.45* |

* $p < .05$.

An efficiency score was calculated for each subject by dividing his performance by his training time. Mean efficiency score as a function of presentation mode and aptitude is presented in Figure 3. Analysis of variance showed that both variables were significant (presentation mode: $F = 19.61$, $df = 3/120$, $p < .001$; aptitude: $F = 15.01$, $df = 1/120$, $p < .001$), and there was no interaction. Specific comparisons showed that each presentation mode was significantly different from the others.

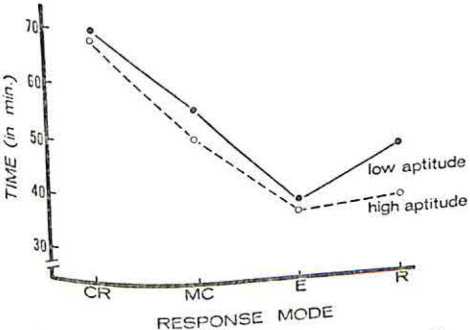


FIG. 2. Training time as a function of response mode and aptitude.

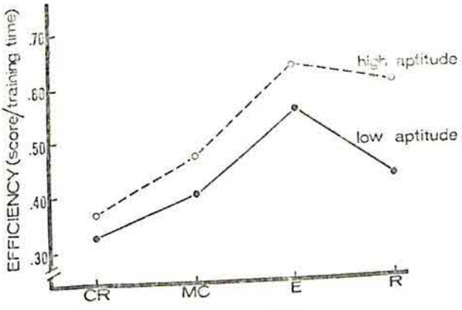


FIG. 3. Efficiency as a function of response mode and aptitude.

These data indicate that Group E was the most efficient mode, especially for low aptitude subjects.

To test the hypothesis that the effectiveness of the different response modes depends on the type of test item, eight items which required that the subject respond with a technical term introduced by the program were selected from the test. Because the distribution of scores was markedly skewed, a square root transformation was made. Figure 4 presents the mean number correct for these items as a function of response mode. An analysis of variance on the transformed scores indicated that there were significant differences among the groups ($F = 2.87$, $df = 3/124$, $p < .05$). The hypothesis under consideration predicts that Group CR will be significantly different from the other three groups, and this specific comparison was made. Indeed, this comparison shows significant differences ($F = 7.73$, $df = 1/124$, $p < .01$). Moreover, there are no differences among the other three groups ($F < 1$).

## DISCUSSION

The general principle that active participation aids learning is supported by the present study. The multiple-choice mode proved as effective as the constructed-response mode in terms of total test score, but these modes were superior to the emphasis mode and the straight reading mode. The fact that differences reached significance in both the present study and that of Holland (1960) may be due to the particular merits of the Holland-Skinner program, which was used in both studies. On the other hand, it may be due to the adequacy of the sample size in both experiments, which was often lacking in other studies.

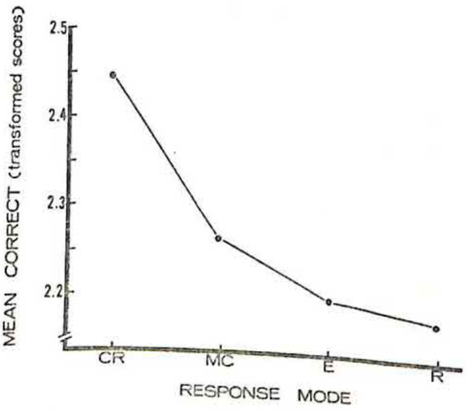


FIG. 4. Score on technical items as a function of response mode.

While Groups CR and MC proved superior in terms of test score, the difference between these groups and the other two was very small. For practical purposes, such small gains might not be worth the greater amount of time taken in those conditions.

Though the performance scores of Groups E and R were comparable, Group R took significantly more time to complete training, and Group E was the most efficient presentation mode. The relative inefficiency of Group R is due to the low aptitude subjects, suggesting that emphasis through underlining can control performance in some cases, and suggesting that this technique be considered in developing learning materials.

Exposure to the program in any form led to scores greatly superior to those of the group which received no training. This difference demonstrates the effectiveness of the program as a review when given some time after original learning. Its effectiveness, of course, may be due to any of several factors, and the present results do not mean that other kinds of review might not prove even more effective.[1]

---

[1] It is interesting to note that even though the program was designed as introductory material, results from a questionnaire administered after the experiment indicated that the subjects (for whom this was the first exposure to programed instruction) tended to favor the

The smaller variance of the CR group and the lack of low scoring subjects in this condition indicate that the program as designed by Skinner does ensure that the minimum level of performance is relatively high. It has previously been shown that achievement scores are more homogeneous for a group of subjects who have used a program than for a group trained by some other method. However, there were many differences between the program group and the control group in these studies, and it is impossible to identify the factors accounting for the differences. In the present experiment only active participation is varied, and it can be concluded that active participation is at least one of the factors which controls the variance of the scores.

The finding that achievement scores were higher for high aptitude subjects than for low aptitude subjects is consistent with that of Lambert, Miller, and Wiley (1962). Yet the question of the relationship of aptitude to the effectiveness of programed instruction cannot be answered on the basis of these findings. Though there *is* a significant correlation between achievement after programed instruction and aptitude (contrary to previous findings), the correlation between performance and aptitude in the group which did not use the program at all was not significantly different. Since the material was review, and since no pretest was given to assess level of performance prior to the experimental treatment, the present findings may be due solely to differences which were present before the program was administered. At any rate, it is clear that differences in proficiency already present cannot be removed by administering a program of this type: after the review, there were still differences in performance as a function of aptitude.

Although a case has been made for the superiority of a constructed-response mode over a multiple-choice mode, there are very few empirical studies which support it. It is suggested here that one reason for the negative findings is that the effects are different for different kinds of items. Test items which required the use of a novel term introduced by the program showed differences between these two response modes. In fact, it is likely that the differences between the two modes were attenuated somewhat in the present study due to the fact that the program was review and thus the terminology already somewhat familiar to the subjects. Moreover, the only other study in which a constructed-response mode and a multiple-choice mode were significantly different did not use a continuous-discourse program (Fry, 1960). Frames in a good continuous-discourse program follow each other logically and closely, and the portion of the frame to be completed by the subject—and the test items—do not necessarily require use of new terminology. Fry's foreign vocabulary program and test, on the other hand, probably included for the most part items which would fall into the category of novel or technical terms.

These conclusions are based on tests administered relatively soon after completion of the program, of course. One possibility which seems not to have been adequately investigated as yet is that though the differences are small in immediate retention, the active participation groups might increase their lead over the others after further instruction of whatever type.

---

use of this type of material in review rather than in original learning. Moreover, in terms of overall attitude, the groups followed the pattern of their performance scores; all were favorable, but Groups CR and MC regarded the program more favorably than did Groups E and R.

## REFERENCES

Coulson, J. E., & Silberman, H. F. Results of an initial experiment in automated teaching. In A. A. Lumsdaine, & R. Glaser (Eds.), *Teaching machines and programmed learning*. Washington, D. C.: National Education Association, 1960. Pp. 452–468.

Evans, J. L., Glaser, R., & Homme, L. E. A preliminary investigation of variation in the properties of verbal learning sequences of the "teaching machine" type. In A. A. Lumsdaine, & R. Glaser (Eds.), *Teaching machines and programmed learning*. Washington, D. C.: National Education Association, 1960. Pp. 446–451.

Ferster, C. B., & Sapon, S. M. An application of recent developments in psychology to the teaching of German. *Harv. educ. Rev.*, 1958, **28**, 58–69.

Fry, E. B. A study of teaching machine response modes. In A. A. Lumsdaine & R. Glaser (Eds.), *Teaching machines and programmed learning*. Washington, D. C.: National Education Association, 1960. Pp. 469–474.

Gates, A. I. Recitation as a factor in memorizing. *Arch. Psychol, N. Y.*, 1917, **6**, No. 40.

Goldbeck, R. A., & Campbell, V. N. The effects of response mode and response difficulty on programed learning. *J. educ. Psychol.*, 1962, **53**, 110–118.

Holland, J. G. Design and use of a teaching machine program. Paper read at American Psychological Association, Chicago, September 1960. (Mimeo)

Hough, J. B. An analysis of the efficiency and effectiveness of selected aspects of machine instruction. *J. educ. Res.*, 1962, **55**, 467–471.

Keislar, F. R., & McNeil, J. D. A comparison of two response modes in an autoinstructional program with children in the primary grades. *J. educ. Psychol.*, 1962, **53**, 127–131.

Krumboltz, J. D., & Weisman, R. G. The effect of overt versus covert responding to programed instruction on immediate and delayed retention. *J. educ. Psychol.*, 1962, **53**, 89–92.

Lambert, P., Miller, D. M., & Wiley, D. E. Experimental folklore and experimentation: The study of programmed learning in the Wauwatosa public schools. *J. educ. Res.*, 1962, **55**, 485–494.

Porter, D. Some effects of year long teaching machine instruction. In E. H. Galanter (Ed.), *Automatic teaching: The state of the art*. New York: Wiley, 1959. Pp. 85–90.

Pressey, S. L. Development and appraisal of devices providing immediate automatic scoring of objective tests and concomitant self-instruction. *J. Psychol.*, 1950, **29**, 417–447.

Skinner, B. F. Teaching machines. *Science*, 1958, **128**, 969–977.

(Received March 15, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 5, 261–267

# INTEREST PATTERNS OF PEACE CORPS VOLUNTEERS IN A TEACHING PROJECT[1]

GEORGE M. GUTHRIE AND MARGARET S. McKENDRY
*Pennsylvania State University*

Analysis of attitude, interest, and adjustment data on 331 Peace Corps volunteers indicates a high incidence of social service motives and esthetic interests. Adjustment patterns on MMPI suggest a lower-than-average incidence of anxiety symptoms and infrequent atypical profiles. Motives for joining include educational, altruistic, and political. Prediction of performance is hazardous due to lack of continuity with previous experience. Many volunteers present a pattern of interests for which only a limited number of satisfying career opportunities exist in our society, since many social service occupations are accorded relatively lower status.

This report is concerned with some measures of personality, interests, and attitudes of a group of Peace Corps trainees. Since the Peace Corps had just been established no adequate information concerning job requirements was available. Since it was a new job, with new people, under new rules, it is not surprising that preliminary results (Hare, 1962) indicate insignificant correlations between predictions of a final American selection board and ratings of performance on the job. Traditionally, prediction of performance is based on samples obtained in various ways of previous performance. This was not possible in this instance because the subjects had not previously performed in an alien culture and had not had any significant previous experience in the role they were to undertake. For these reasons, it would seem important to identify population characteristics for use as potential predictor variables.

## METHOD

### *Subjects*

The subjects were 331 men and women of four groups of volunteers who had been selected for training for service in the Philippines as aides to teachers in rural elementary schools, helping with English as a second language and with beginning science instruction. All but 10 of the volunteers in this sample had completed college. Three-quarters were liberal arts graduates and the balance had been trained as teachers. The majority had completed their education less than a year before joining the Peace Corps. There was one married couple in their fifties and 7 volunteers were between 30 and 40 years of age. The balance were in their twenties with a marked concentration in the age range of 21–24.

### *Procedure*

Tests were administered early in training with the assurance that results were for research purposes and would not be a determining factor in selection. The volunteers were asked to take the MMPI, the Strong Vocational Interest Blank (SVIB), the Study of Values, and a 100-item $Q$ sort of purposes for joining the Peace Corps. This last device was developed by asking groups of college students, not in the Peace Corps, to write down reasons for which they might join. They were classified by inspection, and a set of 100 statements was developed which was presented as reasons that persons had given for joining the Peace Corps. An attempt was made to have reasonably plausible and acceptable statements which would elicit individual differences between volunteers. Examples of statements are "I would like very much to help others," "I want the perspective which this experience would give," and "I believe I have some talents and skills that would be of great value to others." The subject was instructed to sort the statements into a forced normal distribution according to the degree to which each statement expressed his purpose in joining the Peace Corps.

---

[1] This paper was presented at the Annual Meeting of the American Psychological Association, St. Louis, Missouri, September 5, 1962.

## Results

### *MMPI*

The mean MMPI profiles are presented in Table 1 together with data on university student activity leaders reported by Williamson (1952). Scores include the *K* correction. Both men and women volunteers are higher on *K* and lower on *D* and *Pt* than are Williamson's subjects; in addition, male volunteers are higher on the *Mf* scale and lower on *Hs*. Both men and women volunteers were lower, usually significantly lower, in variability on the scales with the exception of the women on the *Hy*, *Mf*, and *Pa* scales. There were few of the profiles which would be judged atypical, a finding which was corroborated by the psychiatric interviews in which fewer than 2% were dropped for psychiatric reasons.

From these data we infer that these subjects are strongly motivated to make a favorable impression. The low *D* and *Pt* scores are indicative of an essentially optimistic outlook with a relative absence of anxiety symptoms. The high *Mf* on males is often found in liberal arts populations and is in keeping with the high esthetic interests of this group which appear on the Study of Values.

### *SVIB*

The results for SVIB are presented in Table 2. The percentage of each group receiving an A or B+ letter score is an indication of their similarity of interest to each occupation. The results are noteworthy in several respects, particularly for the groups of occupations with which many show a high similarity of interests.

The men show high interests in the Group V type of occupations with frequent A scores on the social worker, public administrator, personnel director, and social science teacher scales. They also have high scores on the musician, psychologist, and the physician scales. There is a secondary area of interest for many volunteers in the Group X occupations of journalism and advertising. They are unusually low on most of the

TABLE 1

Mean of MMPI Profiles of Peace Corps Volunteers
(*K* correction included)

| Profile | Male Peace Corps ($N = 168$) | Female Peace Corps ($N = 163$) | Male activity leaders[a] ($N = 467$) | Female activity leaders[a] ($N = 475$) |
|---|---|---|---|---|
| *K* | 63** ± 6.9** | 64** ± 7.0* | 59 ± 8.5 | 58 ± 7.9 |
| *Hs* | 51** ± 7.0 | 49 ± 5.3** | 54 ± 7.5 | 50 ± 6.6 |
| *D* | 49** ± 8.8* | 47** ± 6.8** | 52 ± 10.0 | 48 ± 8.1 |
| *Hy* | 59 ± 6.1* | 56 ± 11.0** | 58 ± 7.0 | 55 ± 7.2 |
| *Pd* | 58 ± 8.4 | 54 ± 8.2* | 58 ± 8.8 | 54 ± 9.3 |
| *Mf* | 66** ± 9.7 | 50 ± 13.2** | 61 ± 9.9 | 49 ± 8.9 |
| *Pa* | 55 ± 7.5 | 55 ± 9.2* | 54 ± 7.5 | 54 ± 7.9 |
| *Pt* | 55** ± 6.6** | 50** ± 5.9** | 57 ± 9.4 | 53 ± 7.6 |
| *Sc* | 58 ± 7.0** | 54 ± 5.7** | 57 ± 8.9 | 55 ± 7.8 |
| *Ma* | 59 ± 8.2* | 57 ± 8.5** | 59 ± 9.7 | 57 ± 10.3 |

[a] Data from Williamson and Hoyt (1952).

* $p < .05$.

** $p < .01$.

TABLE 2

PERCENTAGE DISTRIBUTION OF SVIB SCORES IN THE A AND B+ CATEGORY

| Male Peace Corps volunteers ($N = 167$) | |
|---|---|
| Group occupation | A and B+ |
| I. Artist | 15.6 |
| Psychologist (Rev.) | 55.1 |
| Architect | 28.8 |
| Physician (Rev.) | 47.9 |
| Osteopath | 28.2 |
| Dentist | 3.6 |
| Veterinarian | 1.2 |
| II. Mathematician | 7.2 |
| Physicist | 1.2 |
| Engineer | 6.6 |
| Chemist | 8.4 |
| III. Production Manager | 11.4 |
| IV. Farmer | 12.0 |
| Aviator | 13.8 |
| Carpenter | 2.4 |
| Printer | 31.8 |
| Mathematics or Physical Science Teacher | 28.2 |
| Industrial Arts Teacher | 1.8 |
| Vocational or Agricultural Teacher | 8.4 |
| Policeman | 10.2 |
| Forest Service Man | 15.0 |
| V. YMCA Physical Education Director | 22.8 |
| Personnel Director | 57.5 |
| Public Administrator | 81.5 |
| YMCA Secretary | 36.6 |
| Social Science in High School Teacher | 58.7 |
| City School Superintendent | 42.5 |
| Social Worker | 80.9 |
| Minister | 33.0 |
| VI. Musician (Performer) | 62.9 |
| VII. CPA | 7.2 |
| VIII. Senior CPA | 33.6 |
| Accountant | 7.8 |
| Office man | 16.2 |
| Purchasing Agent | .6 |
| Banker | .0 |
| Mortician | 3.0 |
| Pharmacist | 3.6 |
| IX. Sales Manager | 9.6 |
| Real Estate Salesman | 16.8 |
| Life Insurance Salesman | 31.2 |
| X. Advertising Man | 53.2 |
| Lawyer | 28.8 |
| Author-Journalist | 48.5 |
| XI. President of a Manufacturing Concern | 6.0 |

| Female Peace Corps volunteers ($N = 161$) | |
|---|---|
| Group occupation | A and B+ |
| Artist | 41.0 |
| Author | 33.5 |
| Librarian | 39.1 |
| English Teacher | .0 |
| Social Worker (Rev.) | 79.5 |
| Psychologist | 45.9 |
| Lawyer | 41.0 |
| Social Science Teacher | 13.6 |
| YWCA Secretary | .0 |
| Life Insurance Saleswoman | 1.2 |
| Buyer | .0 |
| Housewife | 1.2 |
| Elementary Teacher | 10.0 |
| Office Worker | 7.4 |
| Stenographer or Secretary | 22.3 |
| Business Education Teacher | 5.6 |
| Home Economics Teacher | .6 |
| Dietitian | .6 |
| Physical Education Teacher | .6 |
| Occupational Therapist | 26.1 |
| Nurse | 9.9 |
| Mathematics or Science Teacher | .6 |
| Dentist | 9.3 |
| Laboratory Technician | .6 |
| Physician | 22.9 |
| Musician Teacher | .0 |
| Musician Performer | 38.5 |

occupations in Groups II, IV, and VIII or the scientific, technical-applied, and business areas.

The women showed patterns most like those of successful social workers, psychologists, lawyers, and artists. A smaller number showed interests which resembled those of librarians, authors, and musical performers. Their interests were particularly unlike those of buyers, dietitians, laboratory technicians, insurance saleswomen, YWCA secretaries, housewives, and the several areas of teaching. These women have a pattern of interests that is infrequent among college students. Although their interests resemble those of social workers, they had not chosen this as a career. It would appear that the blend of patriotism plus service that Peace Corps offers has proven attractive to these women. Their low measured similarity of interest to most teachers suggests that they will not derive their primary satisfaction from their teaching per se.

With both men and women we can assert that there is little similarity of interest to those engaged in sales, business, or science. Rather, both men and women have strong social service motivations, a profound concern for others' welfare. It should be recalled that most of these volunteers had followed traditional liberal arts programs and that not more than a half dozen had given any indication of social work or a cognate field as a career. It is inferred that Peace Corps activity, as they see it, is for many an activity more congenial to their interest pattern than was their college major. From these profiles we would infer that some of these volunteers might enter the service of the Agency for International Development or the State Department, but it does not appear that many will be attracted to a career in business here or abroad.

Further light is shed on these data by the impressions gained from individual interviews. The males of this sample showed a high instance of vocational indecision. Peace Corps experience was an opportunity to express aspects of their interests and at the same time gave them more time to make a career choice. The women, for whom career choice is a different matter, most commonly reported that the Peace Corps was an opportunity to do something significant, an opportunity which was not otherwise open to them. With both men and women service motives, variously expressed, constituted a central theme.

### *Study of Values*

The results of the Allport-Vernon-Lindzey Study of Values are presented in Table 3 together with data on college students reported by Allport, Vernon, and Lindzey (1960). In contrast to MMPI data, the volunteers are more variable than college students on all scales except the political scales; however these differences were significant only in the case of the theoretical scale for males and the political scales for both males and females. Comparisons between volunteers and college students show a significant mean difference on all except the political and religious scales for men and the social scale for women. The practical significance of these differences is difficult to determine because, with samples as large as those involved, small absolute differences become statistically significant. In addition, the method of scoring whereby an individual's total score must sum to a constant causes scores on one scale to automatically affect scores on some other scale, making comparisons difficult.

With these restrictions in mind, the outstanding characteristics of volunteers on this test seemed to be their low scores on economic and their high scores on esthetic interests. Low economic interests would be expected from a group who have volunteered for a low paying service activity. The social scale of this test contains items emphasizing philanthropy and love of people which were not an adequate expression of the social motivations of this group. The high esthetic interest reported is consonant with their liberal education and with their high interests on the musical and verbal linguistic scales of the SVIB, Groups VI and X.

### Q *Sort*

The results of 100 male and of 100 female subjects were analyzed by inverse factor analysis. Four factors for males and three factors for females appeared to be interpretable. Hypothetical person's sorts were generated for these seven factors. These were included with 46 new males and 47 new females to make a third 100 × 100 matrix of persons. The results of this last analysis will be presented here. The factors from the inverse analysis were interpreted by computing the correlation between each item and each factor.

The analysis of the third matrix yielded 10 interpretable factors, which accounted for 61% of the variance. The first 3 of these factors, which by themselves accounted for 37% of the variance, were clearly replications of factors which had been found in the two previous matrices.

The items which correlated most highly with Factor I are shown in Table 4. This factor is a group whose expressed motivations emphasize learning and new experience with a tendency to minimize service to others.

Factor II, with appropriate items shown in Table 5, is made up of individuals whose motives polarize around an assertion of anticommunism and a desire to enhance our own country's prestige abroad. There is a minimization of the extent to which the volunteer himself will be influenced by this experience.

Factor III (refer to Table 6) includes persons who make a strong assertion of personal idealism and altruism and who are not seeking much personal gain from this experience.

The remaining factors were small with very few persons showing high loadings on any single factor. They included motives organized about patriotism, a desire for personal growth, and a drive for novel experience. Further examination of these data indicated that there were no consistent sex differences in responses to this inventory.

An *R* analysis was also performed on

TABLE 3

MEAN STUDY OF VALUES SCORES OF PEACE CORPS VOLUNTEERS COMPARED WITH A COLLEGE SAMPLE

| Scale | Male volunteers ($N = 131$) | Female volunteers ($N = 131$) | College males[a] ($N = 2,489$) | College females[a] ($N = 1,289$) |
|---|---|---|---|---|
| Theoretical | 41.4** ± 8.2* | 40.6** ± 7.5 | 43.8 ± 7.3 | 35.8 ± 7.2 |
| Economic | 32.1** ± 8.6 | 29.9** ± 7.4 | 42.8 ± 7.9 | 37.9 ± 7.3 |
| Aesthetic | 44.8** ± 9.3 | 49.3** ± 9.0 | 35.1 ± 8.5 | 42.7 ± 8.3 |
| Social | 40.8** ± 7.2 | 41.2 ± 7.5 | 37.1 ± 7.0 | 42.0 ± 7.0 |
| Political | 42.8 ± 5.6** | 39.7** ± 5.5* | 43.0 ± 6.6 | 37.8 ± 6.2 |
| Religious | 38.1 ± 10.2 | 39.2** ± 9.9 | 38.2 ± 9.3 | 43.8 ± 9.4 |

[a] From Allport, Vernon, and Lindzey (1960, p. 12).
* $p \leq .05$.
** $p < .01$.

TABLE 4

Items Correlating Highly with Factor 1

| Item number | Item | *r* |
|---|---|---|
| 46 | This experience would help me in my career. | 610 |
| 26 | I want to develop an understanding of other peoples from first hand contact. | 606 |
| 43 | I want to understand world affairs better and I can do so by working in another part of the world. | 595 |
| 62 | Professionally I would benefit greatly from this experience. | 587 |
| 34 | I look on this as a worthwhile opportunity to travel. | 552 |
| 58 | My personal beliefs involve unselfish service to others. | -567 |
| 42 | I have a personal conviction that I should help others. | -529 |
| 22 | I want to lose myself in some unselfish service to others. | -519 |
| 18 | My greatest satisfaction comes from serving others. | -516 |
| 05 | My life is not complete without service. | -463 |

TABLE 5

Items Correlating Highly with Factor 2

| Item number | Item | *r* |
|---|---|---|
| 24 | I want to do my part in defeating communism abroad. | 634 |
| 57 | The Reds are sending people to these countries. I want us to do the same. | 592 |
| 72 | We can defeat communism by proving that we have something better to offer. | 579 |
| 89 | I would like to let the world know the real value of the American way of life. | 550 |
| 41 | I want to help save underdeveloped countries from going communist. | 535 |
| 48 | I will emerge from this with a much clearer idea of the person I want to be. | -554 |
| 75 | I want to be the different person that this service would make possible. | -547 |
| 20 | A change from the things I have been doing would make me a better person. | -487 |
| 11 | I want the perspective that this experience would give. | -461 |
| 28 | This experience would help me become the kind of person I want to be. | -458 |

TABLE 6

Items Correlating Highly with Factor 3

| Item number | Item | *r* |
|---|---|---|
| 73 | This service is an expression of the values which I hold most important. | 729 |
| 97 | This is one of the truest expressions of my belief. | 722 |
| 42 | I have a personal conviction that I should help others. | 584 |
| 58 | My personal beliefs involve unselfish service to others. | 551 |
| 25 | I feel that we have a responsibility to others to help them personally. | 548 |
| 38 | I will be able to do a better job at home after I have had this experience. | -473 |
| 46 | This experience would help me in my career. | -440 |
| 04 | I want to be able to earn the admiration of others. | -425 |
| 70 | I want to help improve the image of the United States in other countries. | -410 |
| 43 | I want to understand world affairs better and I can do so by working in another part of the world. | -400 |

the *Q* sort responses of 330 subjects. Items tended to group around 12 distinct factors: (*a*) professional enhancement versus idealistic altruism, (*b*) improving the image of the United States, (*c*) guiding others to a better way of life, (*d*) desire for travel, (*e*) earning the respect of others, (*f*) finding one's self, (*g*) contacting Americans with overseas experience, (*h*) atoning for mistakes, (*i*) paying a debt owed to the United States, (*j*) helping to defeat communism, (*k*) helping others to help themselves, and (*l*) a willingness to sacrifice in order to serve. Each of these factors is relatively sharply defined. When comparing results of the *R* and *Q* analyses, it appears that people are grouped according to the way in which they respond to combinations of several of the factors from the *R* analysis.

## DISCUSSION

The Peace Corps offers the first opportunity for social scientists to determine, with a reasonably large *N*, factors which influence persons to seek overseas assignments and factors which influence successful performance. The importance of this problem is accentuated by the harm that can be done to international relations by serious failure of this program. The discontinuities of the alien culture with our own cast doubt on traditional prediction and selection variables. Only careful research will determine the degree to which performance overseas is a function of long standing personality and attitude variables and how much it is a function of recent situational and training variables.

Data collected during training on 331 volunteers indicate a high incidence of social service interest patterns on SVIB. However the subjects reject the more philanthropic expression of social values on the Study of Values. The MMPI data suggest a low incidence of symptoms of anxiety. An elevated *K* scale suggests a strong emphasis on appearing symptom free.

Persons with high social service motives and interest in welfare are faced with the problem that our society does not accord high status to many of the occupations in which their interests can be best expressed. In contrast to careers in science or business, careers in social service occupations are not made as attractive or as remunerative. In spite of its low salary scale, the Peace Corps offers these people an opportunity to fulfill some of their most deeply felt purposes, an opportunity which is often not found at home. The impact of this experience on their subsequent careers will warrant careful observation.

## REFERENCES

ALLPORT, G. W., VERNON, P. E., & LINDZEY, G. *Study of values.* (3rd ed.) Boston: Houghton Mifflin, 1960.

HARE, A. P. Factors associated with Peace Corps volunteer success in the Philippines. Haverford: Haverford College, 1962. (Mimeo).

WILLIAMSON, E. G., & HOYT, D. Measured personality characteristics of student leaders. *Educ. psychol. Measmt.*, 1952, **12**, 65–78.

(Received March 18, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 5, 268–276

# SOME NONCOGNITIVE CHARACTERISTICS OF HONORS PROGRAM CANDIDATES[1]

PATRICK J. CAPRETTA, REGINALD L. JONES,[2] LAURENCE SIEGEL, AND LILA C. SIEGEL
*Miami University, Oxford*

Noncognitive personality characteristics of 126 students having had some experience with a university Honors Program were ascertained by means of 3 standard personality inventories and 2 locally devised instruments. As a group, successful Honors students were found to be intellectually oriented toward academic work and rather flexible thinkers; while unsuccessful Honors students and those electing not to participate in the program were decidedly more practical in their approach to college and relatively constricted in thought. No differences were found on other hypothesized characteristics, such as motivation in academic situations and kind of classroom thinking. It is suggested that the "decision" to join is more important than "success" or "failure" in the program in distinguishing among the groups of students.

This study grew out of an attempt to develop an improved set of criteria for selecting Honors Program candidates at Miami University, Oxford, Ohio. The particular program under consideration is typical of other attempts at various colleges and universities to offer an opportunity for students of high academic ability and seriousness of purpose to enroll in special courses and seminars devised to provide a greater challenge to these students than is possible through regular courses alone. Emphasis in such a program is placed upon independent study, and the development of critical thinking and creative capabilities.

Present selection procedures at Miami University consist primarily of "cognitive" predictors, i.e., high school graduating class standing, college course grades (for upperclassmen), and aptitude test scores. Although it has been suggested that school grades do reflect more than the student's ability (Sanford, 1962), little systematic attempt has been made to supplement these objective criteria with information bearing on the so-called "noncognitive" variables, i.e., attitudes, motivation, values, and certain personality characteristics. There is little question as to their general importance in the educational process (Kurland, 1959; Riley, 1959; Warren & Heist, 1960).

The purpose of the present study is to investigate the noncognitive characteristics of various groups of college students having had some association with the Honors Program at Miami University.

## *Hypotheses*

The hypotheses tested in this study stem from preliminary interviews with 13 students representing the extremes of each of three groups. These groups, to be described in more detail later, consisted of students all having had some contact with the Honors Program at Miami University.

Careful study of the interviews revealed what appeared to be a "pattern of approach" to academic work by successful Honors students. These students seemed to be highly motivated by a drive for personal recognition in academic situations. However, this motivation was coupled with a willingness to take a certain amount of personal risk since, by virtue of entering a new program, there was at least a possibility that their grade-point averages might be adversely affected. Furthermore, the interviewer was struck by what seemed to be an inclination within this group to think "divergently," emphasizing implications and generalizations rather than mere factual acquisition. Finally, representatives of this group often voiced their enjoyment and appreciation of opportunities to work independently of the usual classroom structure.

---

[1] The cooperation of the Honors Program Office at Miami University, Oxford, is gratefully acknowledged.

[2] Now at Fisk University.

Thus the overall picture suggested by the interviews was that successful Honor students were, as a group, personally mature, creative, and rather flexible thinkers. In order to examine this pattern more systematically, we induced the following five hypotheses:

1. Successful Honors students (Group AS) will demonstrate greater self-determination and independence of thought than will unsuccessful candidates (Group AU) and those students who decided not to join the Honors Program and who were successful in their non-Honors courses (Group DS).
2. Differences in thinking "style" will be found such that Group AS will appear more fluent and less constricted in thought and belief than will Groups AU and DS.
3. Group AS will be conceptually oriented toward academic work, whereas Groups AU and DS will have a factual set in similar situations.
4. Group AS will value intellectual and esthetic pursuits over and above more "practical" and traditional concerns to a greater extent than will either of the other two groups. (Specific reference is made here to values measured by the Allport-Vernon-Lindzey Study of Values, viz., Theoretical, Aesthetic, Economic, and Religious subscales.)
5. Group AS will demonstrate a higher degree of motivation in academic situations than will Groups AU and DS.

## METHOD

### *Subjects*

One hundred and twenty-six lower division students at Miami University comprised the three groups tested in this study. Definitions of certain terms will be helpful preliminary to listing these groups. *Invitations,* for the individuals under consideration, were given to those students who stood in the upper 10% of their high school graduating classes and received a raw score of 99 or higher on the American College Testing Program Composite. *Success,* for the same students, refers to a 3.25 (A = 4.0) or higher grade-point average made during the students' first semester (or in some cases the first two semesters) in the Honors Program at Miami University.

Group AS consisted of 49 students (17 males and 32 females) who had *accepted* invitations to join the Honors Program and who were *successful* subsequently in their scholastic work. Group AU, on the other hand, comprised 51 students (14 males and 37 females) who had accepted invitations to join the Honors Program but who were *unsuccessful* (i.e., grade-point average below 3.25) in their scholastic work. Group DS consisted of 26 students (10 males and 16 females) who *declined* invitations to join the Honors Program and who were successful in their non-Honors scholastic work.

### *Tests and Measures*

The general research plan was to determine the validity of the foregoing hypotheses by means of various testing instruments which have either been demonstrated in previous research to be applicable to the present study, or appear, on purely logical grounds, to be appropriate measures. The five instruments that were administered are listed below. General descriptive information is offered only for those measures which have been developed locally.

1. Allport-Vernon-Lindzey Study of Values (AVL; 1960). Only four of the six basic values measured by this scale are germane to the present study: Aesthetic (Ae), Economic (Ec), Religious (Re), and Theoretical (Th).
2. California Psychological Inventory (CPI;

Gough, 1957). Only four of the 18 subscores are relevant to this study: Achievement via Independence (Ai), Flexibility (Fx), Psychological-Mindedness (Py), and Self-Acceptance (Sa).

3. Estimate of Time Intervals Inventory (ETI). This inventory, developed by Siegel (1957), purports to be a generalized measure of drive level in terms of expectancies regarding the passage of time. The subject is required to estimate the amount of time that has elapsed between a pair of events. Each item pair consists of a statement of two events which might be expected to transpire successively—e.g., (*a*) Enters college. (*b*) Is recognized by the faculty as an outstanding student. No restrictions are placed upon the response. Subjects may respond in terms of any unit of time measurement they consider to be appropriate. The rationale for ETI is simply that high drive level (to be construed as "motivation" here) is revealed by short-time estimates. The inventory is scored in five homogeneous and relatively independent categories designated: Familial Activities, Heterosexual Activities, Scholastic Activities, Social Activities, and Vocational Activities. Of particular concern for this study is the category of Scholastic Activities. Intended for college students, testing time for the complete ETI is approximately 45 minutes.

4. Omnibus Personality Inventory (OPI). The OPI had its origin as a necessary instrument in a research program at the Center for the Study of Higher Education (1962) in Berkeley. Although in the early developmental stages of this inventory many of the items and scales were taken from reports and surveys of preceding psychological inventories (e.g., MMPI, CPI, VCAI), the major scales in the present OPI represent entirely new combinations of items or highly revised clusters. The OPI is particularly geared to college students and yields 13 subscores. To avoid possible repetition of *kinds* of subtests, only two OPI subscales were given, Complexity (Co) and Thinking Introversion (TI). Testing time for this abridged version of OPI is about 15 minutes.

According to the recently published OPI manual, the complexity measure is related to an "experimental" orientation. High scorers are tolerant of ambiguities, are fond of novel situations and ideas, and are aware of subtle variations in patterns of stimuli, while low scorers require structure and function best in well-organized situations. Persons scoring high on the Thinking Introversion measure are characterized by a liking for reflective thought and display an interest in a variety of ideas for their own sake, whereas the low scorers tend to evaluate ideas on the basis of their practical, immediate application. This last measure appears to reflect a general appreciation of an interest in scholarly activities.

5. Set Toward Education Scale (ESS).[3] This scale, devised specifically for research in higher education by the Instructional Research Service at Miami University, is a forced-choice inventory yielding scores on a continuum characterized on the one hand by a predisposition to learn isolated facts (factual set) and, on the other, by a predisposition to learn principles, relationships, fact clusters, and concepts (conceptual set).

The alternatives constituting each triad were paired on several different preference indexes including judgments about "how bright you have to be to learn this" and "the kind of impression knowing this would make on the naive person." One triad with its scoring weights will serve to illustrate the format of this inventory.

> Items 70–72. Assume you are enrolled in a Natural Science course and must learn about the following. Which one will interest you most? Which one will interest you least?

| | Scoring | |
|---|---|---|
| | Most | Least |
| 70. The names of the elements included with the "halide" group. | −1 | 1 |
| 71. Statement of Newton's Third Law of Motion. | 0 | 0 |
| 72. The significance of a pH of 6. | 1 | −1 |

This scale requires about 15 minutes to administer and yields one score.

Predicted group differences on the various test variables for the five hypotheses are summarized in Table 1.

## *Testing Procedure*

Arrangements for testing as many of the subjects in each of the three groups as possible

---

[3] A copy of this scale has been deposited with the American Documentation Institute. Order Document No. 7692 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance $1.25 for microfilm or $1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

TABLE 1

PREDICTED DIRECTION OF GROUP DIFFERENCES ON THE VARIOUS TEST VARIABLES FOR THE FIVE HYPOTHESES

| Hypothesis | Appropriate test variable | Predicted direction of group differences on test subscales |
|---|---|---|
| 1 | CPI and OPI | AS > AU and DS on Achievement via Independence and Self-Acceptance of CPI and Complexity and Thinking Introversion of OPI |
| 2 | CPI | AS > AU and DS on Flexibility and Psychological-Mindedness |
| 3 | ESS | AS > AU and DS |
| 4 | AVL | AS > AU and DS on Theoretical and Aesthetic<br>AS < AU and DS on Economic and Religious |
| 5 | ETI | AS > AU and DS on Scholastic Activities |

Note.—Group AS—accepted invitation, successful; Group AU—accepted invitation, unsuccessful; Group DS—declined invitation, successful.

were made by sending notices and reminders to each student. From an available pool of approximately 175 subjects, 126 actually participated in the testing program. Of the 49 subjects not tested, there were 8, 21, and 20 from Groups AS, AU, and DS, respectively.

Two to 3 hours were required to complete the test battery, the average time being approximately 2.5 hours. The majority of subjects was tested at one time and remaining ones in small groups spread over a 2-week period. Only the most general instructions were given to the subjects. They were assured that all of the data collected were for research purposes only and would be the confidential property of the Instructional Research Service. Emphasis was placed upon the need for their serious cooperation toward the completion of the research. Most subjects were cooperative and appeared interested in what they were doing. In the few cases for which all of the tests were not completed at one time, the subjects were required to return later (sometimes a week later) to finish the battery.

### *Statistical Analysis*

Major statistical evaluation in this study was accomplished through use of a group-by-sex analysis of variance design. Unequal frequencies in the subclasses necessitated the use of a special procedure described by Walker and Lev (1953). Standard scores rather than raw scores were used for the statistical analysis. The results of this analysis are presented in Table 2.

In light of previous research on sex differences (Freeman & Miles, 1950; Todd, Terrell, & Frank, 1962), analysis by $t$ test for individual groups of males and females treated separately was also made. Probabilities, shown in Table 3, are based upon one-tailed tests.

## RESULTS

### *Hypotheses 1 and 2*

Hypothesis 1, that successful Honors students (Group AS) will demonstrate greater self-determination and independence of thought than will the unsuccessful candidates (Group AU) and those students who decided not to join the Honors Program (Group DS), was largely not supported. No significant differences are shown in Table 2 between the responses of the three groups on the Achievement via Independence and Self-Acceptance scales of CPI or the Thinking Introversion scale of OPI. What little support there is for this hypothesis appears in the comparison between Groups AS and DS on the Complexity scale of OPI ($F=3.98$, $df=1/70$). It is important to note that additional partial support is found for this hypothesis when the results for the male subjects are treated separately (see Table 3). Males in Group AS made significantly higher mean scores on the Achievement via Independence and Self-Acceptance scales of CPI and

TABLE 2—Continued

| Test variable | AS HP successful Male (N = 17) | AS HP successful Female (N = 32) | AU HP unsuccessful Male (N = 14) | AU HP unsuccessful Female (N = 37) | DS Regular successful Male (N = 10) | DS Regular successful Female (N = 16) | Significance level: Between groups AS-AU | Between groups AS-DS | Between groups AU-DS | Between sexes AS-AU | Between sexes AS-DS | Between sexes AU-DS | Group and sex interaction AS-AU | Group and sex interaction AS-DS | Group and sex interaction AU-DS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Religious | | | | | | | | | | | | | | | |
| *M* | 37.2 | 46.0 | 37.3 | 42.9 | 34.8 | 46.2 | *ns* | *ns* | *ns* | .01 | .01 | .01 | *ns* | *ns* | *ns* |
| *SD* | 8.24 | 8.31 | 7.26 | 7.66 | 7.94 | 6.96 | | | | | | | | | |
| Theoretical | | | | | | | | | | | | | | | |
| *M* | 51.8 | 40.1 | 43.9 | 39.4 | 44.3 | 37.6 | .05 | .05 | *ns* | .01 | .01 | .05 | .05 | *ns* | *ns* |
| *SD* | 6.23 | 10.69 | 6.13 | 8.96 | 9.21 | 5.98 | | | | | | | | | |
| ETI | | | | | | | | | | | | | | | |
| Scholastic | | | | | | | | | | | | | | | |
| *M* | 1.25 | 4.00 | .67 | 2.61 | 0 | 2.47 | *ns* | *ns* | *ns* | .01 | .05 | .05 | *ns* | *ns* | *ns* |
| *SD* | 4.63 | 2.98 | 2.59 | 3.54 | 2.40 | 4.59 | | | | | | | | | |

Note.—*M* and *SD* are given in raw score units. Standard scores were used for the statistical analysis. *N* for some of the test variables are somewhat smaller than those given in parentheses.

the Complexity scale of OPI than did males in Groups AU and/or DS. The only significant difference found for females on these same measures is in the opposite direction to that hypothesized (i.e., Self-Acceptance, Groups AS and DS).

Support for Hypothesis 2—that Group AS will be less constricted in thought and belief than Groups AU and DS—is evident in scores on the Flexibility and Psychological-Mindedness scales of CPI. Group AS obtained significantly higher mean scores on these scales than did Group DS (Flexibility: $F = 4.00$, $df = 1/71$; Psychological-Mindedness: $F = 5.01$, $df = 1/71$). In treating the sexes separately by $t$ test, significant differences were found on the Psychological-Mindedness variable for males comparing Groups AS and AU, and Groups AS and DS.

## *Hypothesis 3*

In this hypothesis it is stated that Group AS could be characterized as being conceptually oriented toward academic work, while Groups AU and DS would have a predilection favoring the accumulation of facts in similar situations. In light of the nonsignificant group differences shown in Tables 2 and 3 for ESS, this hypothesis is not supported. However, it should be mentioned that all of the groups considered in the present study have appreciably higher ESS scores (i.e., more conceptually oriented) than do lower division Miami University students in general. The mean ESS score for the latter group is approximately 8, considerably below that of any of the mean ESS scores reported in Table 2. A positive correlation between ESS and scholastic aptitude might help to explain the negative findings with respect to Hypothesis 3.

TABLE 3

RESULTS OF $t$ TEST ANALYSES[a]

| Group | Test variable | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | CPI | | | | OPI | | ESS | AVL | | | | ETI |
| | Ai | Fx | Py | Sa | Co | TI | | Ae | Ec | Re | Th | Scholastic |
| AS and AU | | | | | | | | | | | | |
| Male | 1.72* | *ns* | 2.86** | *ns* | *ns* | *ns* | *ns* | *ns* | *ns* | *ns* | 3.45** | *ns* |
| Female | *ns* | *ns* | *ns* | *ns* | *ns* | *ns* | *ns* | *ns* | 3.15** | *ns* | *ns* | 1.73* |
| AS and DS | | | | | | | | | | | | |
| Male | 2.03* | *ns* | 4.46** | 2.39* | 3.48** | *ns* | *ns* | 1.71* | 3.48** | *ns* | 2.12* | *ns* |
| Female | *ns* | *ns* | *ns* | 2.11[b] | *ns* | *ns* | *ns* | *ns* | 2.89** | *ns* | *ns* | *ns* |
| AU and DS | | | | | | | | | | | | |
| Male | *ns* | 1.94* | *ns* | 1.81* | 2.45** | *ns* | *ns* | *ns* | *ns* | *ns* | *ns* | *ns* |
| Female | *ns* | *ns* | *ns* | *ns* | *ns* | *ns* | *ns* | *ns* | *ns* | *ns* | *ns* | *ns* |

Note.—*df* as follows: AS and AU, males = 29, females = 67; AS and DS, males = 25, females = 46; AU and DS, males = 22.

[a] These $t$'s were computed from $M$ and $SD$ data given in Table 2.

[b] This probability value represents a significant difference (.05) in the direction opposite to that predicted. All other specified probability values are in the predicted direction.

* $p < .05$.

** $p < .01$.

## *Hypothesis 4*

This hypothesis is quite specific in predicting higher Theoretical and Aesthetic scores and lower Economic and Religious scores on AVL for Group AS than for Groups AU and DS. Substantiation of this hypothesis was found in most cases except for the Religious scale. On the Aesthetic scale a difference beyond the .05 level of confidence, shown in Table 2, was found between Groups AS and DS ($F = 5.82$, $df = 1/70$). Significant differences in the predicted direction were also found between the groups in question on Economic (Groups AS and AU: $F = 8.89$, $df = 1/96$; Groups AS and DS: $F = 22.17$, $df = 1/70$) and on Theoretical (Groups AS and AU: $F = 6.61$, $df = 1/96$; Groups AS and DS: $F = 4.82$, $df = 1/70$) subscales. The analysis of variance also indicates appreciable differences between sexes on these scales, especially on the Religious scale (Groups AS and AU: $F = 17.55$, $df = 1/96$; Groups AS and DS: $F = 24.12$, $df = 1/70$; Groups AU and DS: $F = 20.25$, $df = 1/72$), and the Theoretical scale (in the same order given for Religious: $F = 20.31$, $df = 1/96$; $F = 16.90$, $df = 1/70$; $F = 6.23$, $df = 1/72$). For the former AVL scale the females scored consistently higher than did the males, whereas the reverse was true on the Theoretical scale.

The analysis of AVL results by $t$ test, shown in Table 3, provides corroborating evidence that the males in the three groups differ from one another more so than do the females. All four of the significant mean differences are in the predicted direction for the males, while only two are significant and in the right direction for the females.

## *Hypothesis 5*

According to this hypothesis, it was predicted that subjects in Group AS would demonstrate a higher level of motivation than would subjects in Groups AU and DS. Of the overall group comparisons for the Scholastic scale of ETI, shown in Table 2, there are no significant differences, though it should be noted that all of the differences are in the predicted direction. It is also interesting that all of the comparisons between sexes presented in this table resulted in significant differences (Groups AS and AU: $F=8.01$, $df=1/87$; Groups AS and DS: $F=6.30$, $df=1/65$; Groups AU and DS: $F=4.52$, $df=1/63$).

The only significant $t$ test in Table 3 with respect to ETI was found in comparing females in Groups AS and AU, and this difference is in the predicted direction.

## DISCUSSION

The results of this study provide partial support for three of the five hypotheses and thus, in some measure, differentiate successful from unsuccessful Honors Program students and those who decided not to join the program. However, as presented in Table 3, the results differ somewhat depending upon the sex of the subjects, significant differences in the predicted direction being found in 10 of the 24 comparisons for male subjects and in only 3 of the 24 comparisons for female subjects. These results, along with similar findings in other studies (Freeman & Miles, 1950; Todd et al., 1962), serve to emphasize the importance of considering sex differences in such research as represented here.

As might be expected, the significant Sex × Group interactions on test variables Achievement via Independence, Self-Acceptance, Complexity, and Theoretical indicate some degree of similarity among groups of females and dissimilarity among male groups for these variables. In all four cases the successful group of males in the Honors Program scored higher than did the unsuccessful male group and/or those males who decided not to join the program; while little, if any, difference is noted for the three groups of female students.

Both types of statistical analyses disclosed greater discrepancies between the successful Honors and non-Honors groups than between the successful and unsuccessful Honors groups. Table 2 shows significant group differences in the predicted direction on six of the test variables for comparisons between the former pair of groups and only two for the latter pair of groups. Likewise, results presented in Table 3 indicate eight significant $t$'s in the predicted direction for comparisons between the former two and only five significant mean differences for the latter two group comparisons. Also bearing on this relationship are the differences found between the unsuccessful Honors and successful non-Honors groups on test variables Flexibility, Self-Acceptance, and Complexity presented in Table 3. In all cases, males in the unsuccessful Honors group had the higher mean score, i.e., were more flexible, self-confident, etc. In short, students who decided not to join the Honors Program and who were successful in their non-Honors courses differ from those who joined and did well more than this latter group differs from the Honors students who were unsuccessful. The implication here is that "decision" rather than "success" or "failure" is more important in distinguishing among groups of students on the test variables considered in this research. It is unfortunate that the present study did not include a fourth group of the *declined invitation, unsuccessful.*

An interesting parallel is seen between the successful Honors students in the present study and the gifted student in the Warren and Heist (1960) study on gifted college students. This latter research describes the typical gifted student as being strongly attracted to intellectual activity combined with high levels of esthetic awareness and appreciation. Present results on AVL demonstrated that the successful Honors women (as was also true for Warren and Heist's gifted women), in contrast to the men, value the Aesthetic orientation slightly more than the Theoretical, while sharply rejecting the Economic.

The fact that the female subjects made consistently higher motivation scores on the Scholastic scale of ETI than did the male subjects bears upon a hypothesis suggested by McClelland, Atkinson, Clark, and Lowell (1953) that a higher level of scholastic performance is expected of girls than boys at all levels of development. The higher Scholastic score for females is perhaps a manifestation of this expectation. This supposition has some support in the general finding that scholastic underachievement is more frequently found in males than females (Todd et al., 1962).

That personality characteristics other than high academic aptitude do distinguish the intellectually gifted from the general population seems clear from previous research (Warren & Heist, 1960). The results of the present study, to some extent, permit an extension of this finding to include distinctions within the so-called gifted group itself. Not only do noncognitive variables seem to play some role in determining success in an intellectually oriented program but, more important, perhaps, is the difference between superior students who accept invitations to join such programs and those who do not on these same variables.

## REFERENCES

ALLPORT, G. W., VERNON, P. E., & LINDZEY, G. *Study of Values: A scale for measuring the dominant interests in personality.* Boston: Houghton Mifflin, 1960.

CENTER FOR THE STUDY OF HIGHER EDUCATION. *Omnibus Personality Inventory: Research manual.* University of California, Berkeley: CSHE, 1962. (Mimeo)

FREEMAN, F. S., & MILES, C. C. Sex differences. In W. S. Monroe (Ed.), *Encyclopedia of educational research.* New York: Macmillan, 1950.

GOUGH, H. G. *California Psychological Inventory manual.* Palo Alto: Consulting Psychologists Press, 1957.

KURLAND, N. D. The superior failure: Three converging studies. *Super. Student,* 1959, 2(8), 15–18.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

RILEY, S. B. Finding and launching the superior student. *Super. Student,* 1959, 1(7), 9–11.

SANFORD, N. (Ed.) *The American college: A psychological and social interpretation of the higher learning.* New York: Wiley, 1962.

SIEGEL, L. The Estimate of Time Intervals Inventory. Unpublished manuscript, Miami University, Instructional Research Service, 1957.

TODD, F. J., TERRELL, G., & FRANK, C. E. Differences between normal and underachievers of superior ability. *J. appl. Psychol.,* 1962, **46,** 183–190.

WALKER, H. M., & LEV, J. *Statistical inference.* New York: Holt, 1953.

WARREN, J. R., & HEIST, P. A. Personality attributes of gifted college students. *Science,* 1960, **132,** 330–337.

(Received April 1, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 5, 277–285

# VOCATIONAL INTEREST IN RELATION TO VOCATIONAL MOTIVATION

JOHN O. CRITES
*University of Iowa*

This study evaluated the general proposition that vocational motives and interests are related by testing several specific hypotheses about the expected correlations between 7 motivation factors and the occupational scales of SVIB in a sample of 130 male undergraduates. Product-moment r's and analyses of variance supported most of the hypotheses and led to the following conclusions: (a) kind, but not strength, of vocational motivation is related to interests, (b) each interest area on SVIB is associated with a different motivational pattern, and (c) the primary motivational factor which differentiates interest types is the bipolar one of System versus Social Service.

Theories of vocational interest have proposed a number of different hypotheses to account for the well-established phenomenon of individual differences in occupational likes, dislikes, and preferences (Super & Crites, 1962). Empirical tests of the hypotheses, however, with only a few notable exceptions, have yielded largely inconclusive or nonconfirmatory results. Strong's (1943) observation that "interests reflect inborn abilities [p. 682]" has received negligible research support from studies of aptitude and achievement variables in relation to interests, the modal correlation being close to zero (Darley & Hagenah, 1955). Similarly, the relationships of interest to various aspects of personality, such as predominant trait pattern (Darley, 1941), self-concept (Bordin, 1943), and response integration (Carter, 1940), have generally been low (e.g., Cottle, 1950; Tyler, 1945), although sometimes interpreted as theoretically significant (Darley & Hagenah, 1955). Probably the most promising findings on the correlates of interest have come from investigations of family background (Carter, 1932) and parental identification (Crites, 1962) factors, which appear to be related to both the acquisition and patterning of likes and dislikes. Why these associations obtain, however, and what the connecting or linking variables are have not been established.

## PROBLEM

One possible way of conceptualizing and explaining the development of vocational interests, which is consistent with the above research, is suggested by the widely held assumption of practitioners and theorists alike that "interests are motives." In both counseling and personnel selection interests are often interpreted as indicative of level of aspiration, drive, or motivation. Darley and Hagenah (1955) write:

> occupational interests reflect, in the vocabulary of the world of work, the value systems, the needs, and the motivations of individuals [p. 191].

Likewise, Strong (1955) reasons:

> Interest scores measure a complex of liked and disliked activities . . . equivalent to a "condition which supplies stimulation for a particular type of behavior". . . Interest scores are consequently measures of drives.

Drives or motives can, in turn, be related to the social learning conditions of the home as expressed in the attitudes of the parents toward their children and the goals which they value for them, e.g., security, prestige, job freedom. Thus, it may be that motivational variables provide the connection or link between the formation of differential vocational interests and the role of the family in the socialization process. The problem of the present study was to evaluate this conceptual scheme by determining whether a relationship between interests and motives can be demonstrated empirically.

## HYPOTHESES

There has been some research on the relationship of values to interests (Ferguson, Humphreys, & Strong, 1941; Kinnane & Suziedelis, 1962; Sarbin & Berdie, 1940) that can be used in conjunction with recent work on the identification of vocational motivation factors (Crites, 1961) to deduce several hypotheses which are suitable for testing. The hypotheses, and the rationales for them, as formulated for the Strong Vocational Interest Blank (SVIB), are as follows:

1. The desire for Job Freedom, which involves minimal supervision and varied job duties, should be related to interest in SVIB Group I occupations, such as Artist, Architect, Physician, etc., which allow considerable self-direction and are relatively unstructured.

2. The need for System in one's work should find expression in SVIB Group II occupations, such as Mathematician and Engineer, which require method, order, and planfulness in their performance. For essentially the same reason, System should also be related to SVIB Group III and Group IV Technical occupations.

3. The motivations for Behavior Control and Social Service should be associated with the Administrative Contact (Public Administrator, City School Superintendent, etc.) and Personal Contact (Minister, Social Worker, etc.) occupations of SVIB Group V, respectively.

4. The desire for Personal Status, which is an egocentric need for recognition and social position, should be correlated with interests in SVIB Group VI (Musician), Group VII (CPA), and Group X (Verbal) occupations.

5. The need for Structure, or the security which comes from an established routine and stable organizational setting, should be related to interests in SVIB Group VIII (Business Detail) occupations.

6. Finally, the motivation toward Material Security should be related to SVIB Group IX and Group XI (Business Contact) occupations, in which the rewards of work are primarily extrinsic in nature, e.g., commissions, bonuses, visible success.

## METHOD

### *Measuring Instruments*

The vocational motivation variables were measured with questionnaires and self-rating scales which had previously been found to load the highest on the corresponding factorial dimensions defined by several different instruments (Crites, 1961). Material Security, Social Service, and System were assessed by the Materialistic, Humanitarian, and Technical scales, respectively, of the Occupational Attitude Rating Scales (Hammond, 1954); Personal Status, Behavior Control,[1] and Structure were scored from the Status-Need, Managerial-Aggressive, and Organization-Need scales, respectively, of the Work Satisfaction Questionnaire (Astin, 1958); and, Job Freedom was quantified from the Achievement-Need score of the Job Incentive Rankings (Bendig & Stillman, 1958). Typical items from these inventories are illustrated by the following: earn big money, be the Top Man, change people for the better, use a methodological approach, and have good job security. The SVIB was used to measure occupational interests.

---

[1] The Behavior Control factor was originally named Social Approval. Further analysis of its composition, however, has indicated the change in terminology. The present designation was adopted from Caplow (1954).

### *Procedure*

The various measuring instruments were administered to a sample of 130 male undergraduates as part of their experimental participation in an introductory psychology course. To test the hypotheses of the study product-moment correlation coefficients were computed between each of the vocational motivation variables and the 45 occupational and 3 nonoccupational scales of SVIB. The *r*'s for each SVIB group which was comprised of more than one scale were then averaged by *z* transformations. To determine whether strength as well as kind of vocational motivation was related to interests, the subjects with primary, secondary, and reject patterns, as analyzed by Stephenson's (1961) procedures, were compared in each SVIB group on mean motivation scores, using a Type I design (Lindquist, 1953).

## RESULTS

Table 1 presents the data from the correlational analysis of the vocational interest and motivation measures. Approximately 41%, or almost one half, of the *r*'s for the occupational scales and 33%, or about one third, of the *r*'s for the nonoccupational scales were significant at the .05 and .01 levels or beyond. For the separate SVIB groups, the percentages of significant *r*'s ranged from 0 for the single occupational Groups VI and XI to .64 and .62 for the multioccupational Groups II and IX, respectively. The average *r*'s for the interest groups were lowest for the Business Detail, Musician, and President of Manufacturing Concern occupations and highest for the Physical Science and Business Contact fields. The latter areas also had the greatest number of significant *r*'s. The two highest average *r*'s in the table are the .52 between Group II and System and the .42 between Group V and Social Service. With only a few exceptions, the average *r*'s seem to be representative of most of the correlations in SVIB groups, since the standard deviations for the distributions of coefficients within the groups were relatively small, as compared with their possible magnitude, and were uniform from group to group.

Summarizing the *r*'s in Table 1 by reading down the columns, rather than across the rows, indicates that some of the vocational motivation factors are related to more SVIB groups than others. System correlates the most often and the highest with the interest groups, being significantly related to them in 8 out of 11 instances. Social Service and Behavior Control follow with four significant *r*'s for each. In contrast, none of the correlations for Material Security and the interest groups are significant, and only two for Structure are significant (at the .05 level). With respect to the nonoccupational scales, Social Service and Behavior Control correlate positively with Interest Maturity; Material Security and Personal Status relate negatively to Occupational Level; and, Social Service correlates negatively, whereas System is associated positively, with Masculinity-Femininity.

The results from the Type I analyses of variance were negative for the comparisons of subjects within the SVIB groups which had primary, secondary, and reject patterns. Their overall means for the vocational motivation variables, i.e., across the seven factors, were not significantly different in any of the interest groups. As might be expected (McNemar, 1962), however, from the significant *r*'s yielded by the correlational analysis, several of the interactions between SVIB interest patterns and motivation factors were highly significant. The interactions occurred in the same interest groups which had significant correlations, with one notable exception. Whereas none of the *r*'s for Group VIII (Business Detail) was relia-

TABLE 1

PRODUCT-MOMENT CORRELATIONS BETWEEN VOCATIONAL MOTIVATION FACTORS AND SVIB SCALES

| SVIB group | SVIB occupation | Material security | Job freedom | Personal status | Social service | Behavior control | System | Structure |
|---|---|---|---|---|---|---|---|---|
| I. Biological Sciences | Artist | 030 | 230 | −207 | −107 | −340 | 072 | −249 |
| | Psychologist | −157 | 270 | −080 | 173 | 031 | 048 | −102 |
| | Architect | −030 | 270 | −167 | −251 | −275 | 215 | −158 |
| | Physician | −113 | 343 | −247 | −110 | −215 | 351 | −116 |
| | Osteopath | 035 | 241 | −231 | 019 | −138 | 252 | −100 |
| | Dentist | −016 | 258 | −306 | −248 | −324 | 436 | −100 |
| | Veterinarian | 093 | 027 | −188 | −100 | −186 | 217 | −003 |
| | Average *r* | −020 | 240 | −210 | −090 | −210 | 230 | −120 |
| II. Physical Sciences | Mathematician | −167 | 359 | −239 | −203 | −281 | 387 | −070 |
| | Physicist | −089 | 400 | −245 | −288 | −286 | 548 | −019 |
| | Engineer | −014 | 276 | −155 | −353 | −240 | 577 | 033 |
| | Chemist | −006 | 356 | −172 | −273 | −300 | 561 | 016 |
| | Average *r* | −070 | 350 | −210 | −280 | −280 | 520 | −010 |
| III and IV. Technical | Production Manager | 053 | 067 | 073 | −095 | 098 | 353 | 216 |
| | Farmer | 128 | 120 | −047 | −412 | −280 | 349 | 012 |
| | Aviator | 131 | 129 | −014 | −408 | −252 | 364 | −014 |
| | Carpenter | 135 | 169 | −042 | −298 | −231 | 439 | 131 |
| | Printer | 211 | 118 | 034 | −271 | −177 | 185 | 033 |
| | Mathematics and Physical Science Teacher | −019 | 207 | −074 | −068 | −011 | 332 | 129 |
| | Industrial Arts Teacher | 032 | 188 | −155 | −238 | −160 | 383 | 096 |
| | Vocational and Agricultural Teacher | 022 | −001 | −096 | −121 | −077 | 152 | 065 |
| | Policeman | 147 | −047 | 084 | −051 | 056 | 163 | 174 |
| | Forest Service Man | 017 | 075 | −058 | −200 | −046 | 280 | 060 |
| | Average *r* | 090 | 120 | −040 | −230 | −130 | 290 | 080 |
| V. Social Welfare | YMCA Physical Director | 003 | −067 | 003 | 357 | 317 | −116 | 066 |
| | Personnel Director | −059 | −132 | 182 | 328 | 427 | −295 | 078 |
| | Public Administrator | −166 | 029 | 116 | 357 | 411 | −201 | 100 |
| | YMCA Secretary | −127 | −101 | 106 | 466 | 413 | −350 | 078 |
| | Social Science Teacher | 008 | −179 | 173 | 450 | 357 | −413 | 071 |
| | City School Superintendent | −180 | −020 | −002 | 449 | 363 | −264 | 021 |
| | Social Worker | −100 | −043 | 106 | 470 | 364 | −307 | 040 |
| | Minister | −132 | 061 | −035 | 419 | 248 | −233 | −043 |
| | Average *r* | −100 | −060 | 080 | 420 | 360 | −290 | 050 |
| VI, VII, and VIII. Business Detail | Musician | 056 | 100 | −059 | 128 | −042 | −030 | −124 |
| | CPA | −161 | 036 | 050 | 181 | 215 | −087 | 050 |
| | Senior CPA | −016 | −023 | 091 | −124 | 132 | 048 | 130 |
| | Accountant | 013 | −046 | 101 | 013 | 218 | 020 | 205 |

TABLE 1—Continued

| SVIB group | SVIB occupation | Material security | Job freedom | Personal status | Social service | Behavior control | System | Structure |
|---|---|---|---|---|---|---|---|---|
| | Office Man | 159 | −200 | 212 | 071 | 224 | −230 | 200 |
| | Purchasing Agent | 140 | −108 | 212 | −165 | 023 | −049 | 100 |
| | Banker | 033 | −321 | 191 | 029 | 218 | −211 | 212 |
| | Mortician | 136 | −286 | 148 | 207 | 155 | −242 | 149 |
| | Pharmacist | 251 | −080 | 086 | 065 | −107 | −007 | 023 |
| | Average *r* | 100 | −150 | 150 | 010 | 120 | −100 | 140 |
| IX. Business Contact | Sales Manager | 037 | −265 | 248 | 239 | 335 | −364 | 010 |
| | Real Estate Salesman | 133 | −341 | 290 | 166 | 094 | −382 | 008 |
| | Life Insurance Salesman | 025 | −306 | 179 | 452 | 321 | −494 | 019 |
| | Average *r* | 060 | −310 | 240 | 290 | 250 | −400 | 140 |
| X. Verbal | Advertising Man | 040 | −082 | 134 | 163 | 068 | −493 | −188 |
| | Lawyer | −047 | −088 | 057 | 268 | 153 | −398 | −162 |
| | Author-Journalist | 037 | 099 | −062 | 005 | −212 | −207 | −294 |
| | Average *r* | 010 | −020 | 040 | 150 | 000 | −370 | −210 |
| XI. President of a manufacturing concern | | −008 | 053 | 023 | 053 | 041 | 169 | 090 |
| Nonoccupational scales | Interest Maturity | −148 | −039 | 051 | 239 | 296 | −219 | 036 |
| | Occupational Level | −188 | 059 | −225 | 103 | 033 | −045 | −106 |
| | Masculinity-Femininity | −009 | 065 | −035 | −419 | −133 | 382 | 087 |

Note.—*r*'s of .174 and .228 are significant at the .05 and .01 levels, respectively. Decimal points have been omitted.
$N = 130$.

ble, the $F$ (4.88, $df = 6/762$) for the interaction was significant at the .001 level. Further analysis of this interaction with $t$ tests revealed that it arose from differences in the means for the Job Freedom and Personal Status factors. The subjects who had primary or secondary patterns in the Business Detail group scored relatively higher on motivation for Personal Status as compared with Job Freedom than the subjects who had reject patterns in this interest area. The means for these two groups of subjects on the factors and the $t$ value for the interaction, which was significant at the .001 level, were as follows: Primary/Secondary—Personal Status = 19.77, Job Freedom = 8.33; Reject—Personal Status = 16.85, Job Freedom = 10.32; $t = 3.29$.

## DISCUSSION

In general, the results were favorable to the hypotheses, indicating that kind (but not strength) of vocational motivation is related to interests, but there were some noteworthy exceptions. Some predicted relationships were not obtained, and some obtained relationships were not predicted, as the following discussion of the correlations for each SVIB group points out:

*Group I (Biological Sciences).* Job Freedom correlated with interests in this group, as hypothesized, but so did System to about the same extent. Also, there were negative correlations with Personal Status and Behavior Control. Apparently, the individual with interests in this group is motivated by a desire for autonomy in his work, as expressed in being on his own as well as *not* controlling the activities of others, and by a need for an orderly and planful approach to his work. It may be that it is this motivational complex which accounts for the high intercorrelations of the occupational scales in this group, which are otherwise difficult to rationalize. If so, it may make better sense for counselors and personnel workers to interpret Group I interests in terms of common motives rather than job duties and tasks, and thus resolve the worrisome problem of including Artist and Architect in the Biological Sciences.

*Group II* (*Physical Sciences*). The predicted relationship of System to interests in this group was strongly supported by an *r* of .52. As in Group I, other correlations were also significant, notably the positive relationship with Job Freedom and the negative associations with Personal Status, Social Service, and Behavior Control. Taken together, these relationships seem to characterize Physical Science interests as they have usually been interpreted: a focus upon working with things methodically and systematically in a relatively free situation with little emphasis upon personal aggrandizement.

*Groups III and IV* (*Production Manager—Technical*). The hypothesis that System would be related to these occupations was substantiated. In addition, Structure was positively related to Group III and Social Service was negatively correlated with Group IV. The pattern of *r*'s for these two groups is similar to that for Group II, with the important difference that there is no relationship with Job Freedom. Since the Technical occupations are at a lower level of responsibility and skill than the scientific professions, however, the difference in motivation for autonomy in one's work might be expected.

*Group V* (*Social Welfare*). Again, the predictions were upheld, with Social Service and Behavior Control correlating with interests in this group. There was no differentiation, however, between these factors in their relationships to the Administrative and Personal Contact occupations. The negative correlation with System is explicable; since this motive relates to working with things rather than with people. Thus, the usual interpretation of Group V interests as socially oriented is consistent with the motivational factors which are related to them.

*Groups VI, VII and XI* (*Musician, CPA, President*). The correlations for these single-occupational groups were generally low and nonsignificant and did not support the predictions. None of the *r*'s for Musician was large enough to be significant, and those for CPA and President were contrary to expectation. As a result, little can be said about the motivations which are related to interests in these occupations other than to note that further study of them is needed.

*Group VIII* (*Business Detail*). The hypothesized relationship of Structure to this group was not confirmed by either the correlational or the Type I analyses, although the latter indicated that there is a motivational pattern which may be related to Business Detail interests. There appears to be a tendency for individuals with primaries and secondaries in this area to have a motivation toward Personal Status, but this becomes apparent only when considered in conjunction with their motivation for Job Freedom, which is relatively less strong. Because the correlational analysis dealt with the relationship of only one motivational variable to interests at a time, instead of all factors simultaneously as in the Type I design, it did not reveal this pattern. It makes considerable post hoc sense, however, because of the prestige and status which can be gained through white-collar, office-type work.

*Group IX* (*Business Contact*). Contrary to prediction, interests in this group were related to practically every other motivational factor except Material Security. There were positive correlations with Personal Status, Social Service, and Behavior Control and negative associations with Job Freedom and System, all of which suggest that sales interests reflect a complicated motivational system consisting of egocentric as well as social welfare and possibly dependency needs. The general stereotype which emerges of the salesman is that of a fairly narcissistic individual who needs others to enhance his self-image but who also strives to achieve social goals and objectives.

*Group X* (*Verbal*). The correlations for this group are somewhat difficult to explain. The hypothesized Personal Status relationship was not obtained, but negative associations with System and Structure were. Without any positive correlations for the other motivation factors, the meaning of these relationships is not clear, but they may reflect an antagonism toward authority, as implied by rebellion against System and Structure, which is often characteristic of individuals with verbal interests.

*Nonoccupational Scales*. The correlations with the nonoccupational scales generally agree with their usual interpretations. Interest Maturity was positively related to motivations for working with people (Social Service, Behavior Control); Occupational Level was negatively associated with sources of extrinsic satisfactions in work (Material Security, Personal Status); and, Masculinity-Femininity of interests was positively correlated with a nonperson (System) and negatively related to a person (Social Service) orientation.

The motivation factors which accounted for the largest amount of variation in interests across the 11 SVIB groups were the System, Social Service, and Behavior Control variables. In combination, these motives seem to correspond rather closely to the "Things versus People" bipolar factor identified by Thurstone (1931) and Strong (1943) in factor analyses of SVIB, which suggests that this interest-motivation dimension may be the basic one along which occupations are differentiated. Job Freedom and Personal Status make some contribution to individual differences in interests, but not as much as the other factors. Material Security and Structure are essentially unrelated to any of the interest groups, which agrees with the common interpretation of SVIB as a measure of "intrinsic and self-perpetuating satisfactions" rather than the extrinsic and concomitant rewards of work (Darley & Hagenah, 1955, p. 12). The studies by Sarbin and Berdie (1940) and Kinnane and Suziedelis (1962) would generally be consistent with these conclusions and increase our confidence in them. Theoretically, the identification of a person-nonperson orientation which underlies differential interest patterning fits Roe's (1957) hypotheses about the relationships of these variables to each other and to parental attitudes which may have influenced early vocational behavior. The conceptualization of vocational motivation as an intervening variable between familial conditions and interest development would appear to have some promise for both explanatory and heuristic purposes.

Alternative interpretations of the results should be considered, however, since it might be argued that the correlations of the interest and motivation measures may reflect only common test variance or represent simply covariation between two associative tendencies. A motivational variable is usually defined as having two characteristics: (*a*) it has the associative or cognitive property of facilitating the occurrence of specific responses, and (*b*) it acts in a more general way, i.e., to energize indiscriminately all of the reaction tendencies in a given situation and/or to lead to the learning of responses which eliminate or reduce its intensity (Farber, 1955; Melton, 1950; Shaffer & Shoben, 1956). Although very few variables have been shown to possess these last two attributes, they are the critical ones in distinguishing a variable as motivational (Brown, 1961). It is important to ask, therefore, whether the so-called vocational motivation factors function so as to either stimulate a variety of behaviors or produce learning. What evidence there is on this problem is admittedly sketchy and open to many of the criticisms which have been made of other test definitions of motivation (Child, 1954; Farber, 1955), but it nevertheless provides a presumptive basis for inferring that the instruments used in this study have some validity as measures of motivational variables. The most relevant data come from a study by Hammond (1959) in which she found that survivors of a 5-year college engineering curriculum had scores on the Occupational Attitude Rating Scales which conformed more closely to the faculty's perception of the "ideal" student at the end of the training program than at the beginning. In other words, this instrument was related to persistence in an activity over a long period of time and acquisition of responses appropriate to the attainment of a desired goal, both of which are indicative of the operation of a motivational variable. Additional research is definitely needed, but it appears justifiable to conclude tentatively that vocational motives can be validly assessed with inventories and questionnaires and that they are systematically related to vocational interests in many of the ways which current theories propose.

## REFERENCES

ASTIN, A. W. Dimensions of work satisfaction in the occupational choices of college freshmen. *J. appl. Psychol.*, 1958, **42**, 187–190.

BENDIG, A. W., & STILLMAN, EUGENIA L. Dimensions of job incentives among college students. *J. appl. Psychol.*, 1958, **42**, 367–371.

BORDIN, E. S. A theory of vocational interests as dynamic phenomena. *Educ. psychol. Measmt.*, 1943, **3**, 49–66.

BROWN, J. S. *The motivation of behavior.* New York: McGraw-Hill, 1961.

CAPLOW, T. *The sociology of work.* Minneapolis: Univer. Minnesota Press, 1954.

CARTER, H. D. Twin similarities in occupational interests. *J. educ. Psychol.*, 1932, **23**, 641–655.

CARTER, H. D. The development of vocational attitudes. *J. consult. Psychol.*, 1940, **4**, 185–191.

CHILD, I. L. Personality. *Annu. Rev. Psychol.*, 1954, **5**, 149–171.

COTTLE, W. C. Relationships among selected personality and interest inventories. *Occupations*, 1950, **28**, 306–310.

CRITES, J. O. Factor analytic definitions of vocational motivation. J. *appl. Psychol.*, 1961, **45**, 330–337.

CRITES, J. O. Parental identification in relation to vocational interest development. *J. educ. Psychol.*, 1962, **53**, 262–270.

DARLEY, J. G. *Clinical aspects and interpretation of the Strong Vocational Interest Blank.* New York: Psychological Corporation, 1941.

DARLEY, J. G., & HAGENAH, THEDA. *Vocational interest measurement.* Minneapolis: Univer. Minnesota Press, 1955.

FARBER, I. E. The role of motivation in verbal learning and performance. *Psychol. Bull.*, 1955, **52**, 311–327.

FERGUSON, L. W., HUMPHREYS, L. G., & STRONG, F. W. A factorial analysis of interests and values J. *educ., Psychol.*, 1941 **32**, 197–204.

HAMMOND, MARJORIE. Occupational attitude rating scales. *Personnel Guid. J.*, 1954, 32, 470–474.

HAMMOND, MARJORIE. Attitudinal changes of "successful" students in a college of engineering,. *J. counsel. Psychol.*, 1959, 6, 69–71.

KINNANE, J. F., & SUZIEDELIS, A. Work value orientation and inventoried interests. *J. counsel. Psychol.*, 1962, 9, 144–148.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

MELTON, A. W. Learning. In W. S. Monroe (Ed.), *Encyclopedia of educational research.* (2nd ed.) New York: Macmillan, 1950. Pp. 668–690.

MCNEMAR, Q. *Psychological statistics.* (3rd ed.) New York: Wiley, 1962.

ROE, ANNE. Early determinants of vocational choice. *J. counsel. Psychol.*, 1957, 4, 212–217.

SARBIN, T. R., & BERDIE, R. F. Relation of measured interests to the Allport-Vernon study of values. J. *appl. Psychol.*, 1940, 24, 287–296.

SHAFFER, L. F., & SHOBEN, E. J., JR. *The psychology of adjustment.* (2nd ed.) Boston: Houghton Mifflin, 1956.

STEPHENSON, R. R. A new pattern analysis technique for the SVIB. *J. counsel. Psychol.*, 1961, 8, 355–362.

STRONG, E. K., JR. *Vocational interests of men and women.* Palo Alto: Stanford Univer. Press, 1943.

SUPER, D. E., & CRITES, J. O. *Appraising vocational fitness.* (Rev. ed.) New York: Harper, 1962.

THURSTONE, L. L. A multiple-factor study of vocational interests. *Personnel J.*, 1931, 10, 198–205.

TYLER, LEONA E. Relationships between Strong Vocational Interest scores and other attitude and personality factors. *J. appl. Psychol.*, 1945, 29, 58–67.

(Received April 3, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 6, 287–294

# DIVERGENT CONCEPTS OF TEACHER ROLE:
## AN APPROACH TO THE MEASUREMENT OF TEACHER EFFECTIVENESS[1]

A. GARTH SORENSON, T. R. HUSEK, AND CONSTANCE YU
*University of California, Los Angeles*

Role theory offers a potentially important approach to the investigation of teacher behavior and effectiveness. This paper reports the 1st steps in the development of an instrument designed to assess teacher role expectations. A preliminary form administered to 284 students was designed to measure 6 role dimensions: information giver, disciplinarian, advisor, counselor, motivator, and referrer. A factor analysis confirmed these 6 factors, and provided a basis for a revised form which was administered to 94 students. These results also were factor analyzed. On the basis of the 2 studies, a revised set of keys was developed for 5 of the 6 scales; advisor and information giver were combined to form a single scale. Reliability estimates for the scales range from .77 to .93. The scales have low to moderate intercorrelations.

A major obstacle to developing valid procedures for predicting teacher effectiveness has been lack of a solution to the criterion problem. Investigations have shown with discouraging frequency that any two observers, whether or not in the educational field, are likely to disagree sharply in the appraisal of a particular teacher. The crux of the disagreement is pinpointed in the fact that any concept of teacher effectiveness is inevitably relative. As Ryans (1960) noted in *Characteristics of Teachers,*

> Two very important reasons why effective and ineffective teachers cannot be described with any assurance are the wide variation in the value concepts underlying descriptions of desirable teaching objectives and the differences in teacher role at different educational levels, in different subjects and with different pupils [p. 371].

If two individuals, in speaking of the "good teacher," are in fact likely to have in mind different traits because they expect him to perform different functions, then they are also likely to disagree when asked to judge a particular teacher because he may satisfy the expectations of the one but not the other. It is erroneous to assume that terms like "good teacher" or "effective teacher" have a common meaning. It is impossible to talk about teacher effectiveness except within a particular system of values and a set of expectations based on those values. The first step in predicting teacher effectiveness, therefore, should be the development of a measure to identify and study different expectations. This paper presents the initial stages in a research program designed to examine the nature and influence of teacher role expectations.

The concept of role expectation as used in this approach is similar to Sarbin's (1954). A society may be regarded as structured into a number of positions, statuses, or offices. Each of the positions, designated by a single term, such as father, policeman, or teacher, is a collection of rights and duties. The actions of the individual are organized around these positions and comprise the role. The role is linked with a position and not with the person who occupies the position. To the extent that the actions of a person in a particular position are perceived as conforming to the expectations (preferences) of the members of his society, he may be said to be successful in fulfilling that role. To the extent that he fails to behave in the prescribed manner, that he does not conform to the role expectations, he is subject to censure.

---

[1] The authors would like to express their appreciation to the Western Data Processing Center, University of California, Los Angeles, for the use of facilities in processing the data reported in this study and to Shirley Schaefer for supervising the analysis of data.

In our society there is no universal set of expectations which defines a teacher's role. Instead, an unknown number of sets is held by people from varying cultural and socioeconomic backgrounds, each individual being further differentiated by his particular experience and specific relationship to the teaching profession. In order to predict how a particular observer will judge a teacher, it is necessary first to discover what the observer's expectations are with respect to the teacher's role. It is essential to determine the criteria he employs in judging and the actions or qualities pertinent to the formation of his opinion of a teacher.

Knowledge of the individual's role expectations cannot be obtained by the simple expedient of asking him about his criteria, as he is probably unaware of some of the most crucial factors. He is unable to state precisely the determinants of his judgment. If he is asked about the basis for his judgment of teachers, what he is likely to report is a series of abstractions—intelligent, friendly, and efficient—abstractions for which there are no easily applied behavioral definitions. As indicated by Sarbin (1954), role expectations are cognitive and must be inferred either from phenomenological report or from the analysis of other kinds of data.

Described below is an instrument from which it is hoped inferences can be made regarding teacher role expectations, specifically with regard to interpersonal relations. Among the hypotheses which might be tested with such an instrument are the following:

1. When teachers and students are paired according to role expectations, greater satisfaction will be experienced by both teachers and students.
2. The effectiveness of a teacher may be predicted by discovering the role expectations of the teacher and his students.
3. The teaching behavior of a teacher or teacher candidate may be predicted from his role expectations.
4. If the role expectations of an individual are judged to be inappropriate, they may be changed deliberately, though in some instances probably only with considerable difficulty, since roles are learned.

The several role dimensions utilized in the first form of the questionnaire were developed in the following manner. In discussions with teachers, administrators, and others interested in teacher training, each individual was asked to think of one of the best teachers he knew and also one of the poorest, and then to describe the essential differences between the two. After approximately 2 dozen such interviews, the notes made during the interviews were reviewed and an attempt was made to categorize the descriptive statements, each category to represent a role dimension. During this stage of the research, various articles relating to teaching behavior and teacher evaluation were examined in a search for ideas which would be of use in further clarifying the emerging role dimensions. A paper by Sears (1957) was especially helpful. Six teacher roles were postulated, as follows:

1. Advisor. In this role the teacher uses his own experience and knowledge

to recommend strongly a particular course of action for the student; he solves the problem and decides for the student.

2. Counselor. The teacher who assumes this role seeks basic causes underlying behavior; concerns himself primarily with the student, helps him discover more courses of action on which to base his decision; wants the student to think independently.

3. Disciplinarian. This teacher adheres rather rigidly to rules, authority, and tradition and believes in punishment as an incentive.

4. Information giver. In this role the teacher gives information, but does not indicate approval or disapproval; he regards the directing of learning as his main function.

5. Motivator. The teacher manipulates situations so that the student will be stimulated to action predetermined by the teacher; he uses rewards, implied or real, as incentives.

6. Referrer. In some circumstances, instead of dealing directly with the problem himself, the teacher secures the help of various agencies available: counselor, principal, special classes, and so on; he shifts the responsibility and the problem to others.

In the course of performing his duties, depending on the circumstances, any teacher probably will assume all of the above roles at one time or another; he will shift from information giver to disciplinarian to advisor and so on. Nevertheless, it is hypothesized that individual teachers will differ in their tendencies to take one or more roles in preference to the others and that these tendencies are relatively stable characteristics of the individual. If these hypotheses are sound, then it should be possible to construct an instrument which, by allowing a subject to select among possible role behaviors, will provide a picture of the pattern of his tendencies and also make it possible to predict his future behavior.

## QUESTIONNAIRE

The Teacher Practices Questionnaire (TPQ) consists of 30 problem situations typical of those encountered by teachers in their daily routine. Each problem situation is a brief description of a student and his behavior. For example:

> Jerry, IQ of 140, enrolls in as many "snap" courses as possible in high school, and receives B's and C's with little effort. He tells his schoolmates, "Get wise. Don't kill yourself with homework. The school will have to graduate you when you are of age anyway."

In constructing the problem situations, an attempt was made to provide simple and realistic descriptions and to avoid as much as possible the use of value-laden words which might influence the respondent's behavior. The object was to give the respondents the opportunity to react to the problem situation per se and not to be biased by emotional terms.

For each problem, four alternative solutions were presented representing different role dimensions. Each of the possible 15 different combinations of roles was presented twice, each of the six roles being represented 20 times in the instrument.

For the example given above, the following alternatives were presented:

1. Give him a good "talking to."

2. Investigate the possibility of an emotional block.

3. Refer him to the school psychologist.

4. Tell him that he should make wiser use of his native ability and lead his classmates accordingly.

The subjects were instructed to rate each of the alternative courses of action on a 5-step scale as to the degree of appropriateness:

1. Very appropriate
2. Fairly appropriate
3. Acceptable
4. Fairly inappropriate
5. Very inappropriate

Each alternative was to be scored 1, 2, 3, 4, or 5, depending on the choice of the respondent, and thus there was a possible range of scores from 20 to 100 for each role dimension.

## STUDY 1

### *Subjects*

The subjects for Study 1 were 284 students in the School of Education at the University of California, Los Angeles, working for teaching credentials. One hundred and twenty-eight were candidates for the elementary credential, and 156 were candidates for the secondary credential; there were 64 men and 224 women. These students were divided into the following subsamples: (*a*) 64 beginning secondary credential candidates, men; (*b*) 94 beginning secondary credential candidates, women; (*c*) 126 beginning elementary candidates, women. All of these subjects completed the questionnaire during the fall semester of 1961, their first semester in the School of Education. Approximately 40% of the students were juniors, 40% seniors, and 20% graduate students.

### *Results*[2]

For each subsample, split-half reliability estimates were obtained for each of the role dimensions as well as intercorrelations between the roles. The split-half reliabilities ranged between .75 and .92. The intercorrelations of scales ranged between −.18 and .80 with most of the correlation coefficients between .30 and .50. The reliabilities were deemed adequate for a preliminary form of an instrument, but the intercorrelations between the scales were far too high. Consequently it was decided to perform a factor analysis and to use the results of the factor analysis and theory underlying the instrument to revise the questionnaire. The three subsamples were combined into one group for factor analysis.

A principal-components analysis was performed on the data from the 120 items and three separate rotations were also performed. Kaiser's varimax criterion was used for rotations in the spaces defined by the four, six, and eight largest factors. Three separate rotations were computed because analytic rotations are somewhat sensitive to the number of factors that are rotated. The results when six factors were rotated were the most reasonable, and these results were used to assist in the revision of the instrument. Five of the factors were interpreted as reflecting five of the role dimensions. The motivator scale, however, did not appear to be reflected clearly in any single factor. In addition, although the other five role dimensions were clearly delineated or embodied in the analysis, each having its own factor, many of the individual items did not load highly on the factors that we would have predicted.

At this point, a decision had to be made whether to relocate the items on the basis of the factor analysis, which would have resulted in a questionnaire problem situation having more than one item scored on a single role dimension, or to rewrite the items and try to change their factorial composition.

---

[2] Many of the results presented and discussed in this paper would be more meaningful if more data were presented. However, space limitations prohibit this. Consequently a 16-page document containing a complete copy of the revised form of the questionnaire, a listing of the original scales, and the five revised scales, and for each of the 120 items in Form 2, means, standard deviations, and loadings on the six rotated factors obtained from Study 2 have been deposited with the American Documentation Institute. Order Document No. 7707 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance $1.75 for microfilm or $2.50 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

We chose to rewrite the items and attempt to obtain six role dimensions of the kind originally postulated. Of the 120 items, 51 were rewritten for the second version of the questionnaire.

## STUDY 2

### *Subjects*

Subjects for Study 2 were 94 students in the School of Education at the University of California, Los Angeles, working for the secondary teaching credentials. Sixty were women and 34 were men. These students were enrolled in a course in educational psychology required of teacher candidates. Most of the students in this sample were graduates and about half had 1 or more years of teaching experience.

### *Results*[2]

Means, standard deviations, and intercorrelations of the 120 items in the revised form were computed. The item intercorrelations were subjected to a principal factor component analysis and, just as in Study 1, three different rotations were performed using Kaiser's varimax criterion. This time, however, the patterning of the unrotated factors led us to rotate four, six, and nine factors.

As with the first study, the rotation using six factors was the most reasonable. The patterns of the loadings were quite consistent with the results obtained for Study 1, and four of the factors could be clearly identified as reflecting the disciplinarian, counselor, motivator, and referrer role dimensions. The postulated dimensions called advice giver and information giver seemed to collapse.

The results of the factor analyses for Study 1 and for Study 2, along with considerations arising out of the theory underlying the instrument, were used to generate five revised keys for the five dimensions of disciplinarian, counselor, motivator, referrer, and advice-information giver (see Table 1).

The criteria used were as follows: If an item had been revised for Form 2, then the size of the loading of the item on a factor from the second study (and to some extent the size of the loadings on the other factors), together with our judgment as to the appropriateness of the item for the factor in terms of the underlying theory for the test, were used in evaluating and locating the item on a scale. If an item had not been revised, then we used both factor analyses and the theory in our revision. Although our manner of using the factor analytic results was not always the same from item to item, in virtually no case was an item placed on a key on which its loading was less than .40. Since we could not always find 20 items for a particular key, some of the revised scales have less than 20 items. It should also be mentioned that, although we selected items for the keys which possessed large loading on the appropriate factors and low loading on the other factors, we made no attempt to obtain keys which would provide uncorrelated estimates of the factors. The matter of relationships among the various scales will be explicated in the discussion. Means, standard deviations, and split-half reliability estimates for these new scales were obtained and are presented in Table 2.

The intercorrelations between these new scales were also obtained (see Table 3). It should be noted that the intercorrelations have been obtained from data which were used in part to obtain these scales. These correlations must be interpreted with some caution,

TABLE 1

EXAMPLES OF PROBLEMS AND ITEMS FOR THE REVISED SCALES OF TPQ, FORM 2

| Problem | Item | Scale |
|---|---|---|
| III. A sophomore student in engineering is uncertain about his vocation. He comes to you, a professor of engineering, for advice. He likes his courses and is making good grades. He believes that it will be financially profitable to be an engineer; however, he is also strongly interested in English. He is altruistic and wonders whether he can perform a greater service in the field of English. | 9. Provide him with information about pay scales, opportunities for placement and advancement in both professions. | M |
| | 10. Tell him that there is great need for good engineers in this Space Age, that he will find a career in engineering most challenging and worthwhile. | AI |
| | 11. Attempt to help him analyze the causes underlying the uncertainty. | C |
| IX. During the past few weeks, the fourth graders have complained about losing pencils and erasers. Mary is discovered to be the culprit. She comes from a middle class family and is provided with more spending money than the average student. | 33. Explain to Mary why the other students are annoyed about losing their possessions. | M |
| | 34. Tell her parents about her behavior, and let them deal with the problem as they see fit. | R |
| | 35. Warn Mary that she will be suspended the next time she steals. | D |
| XIII. Michael is 13. His parents, itinerant farm laborers, have been forced by the juvenile authorities to keep him in school. He seems to be placed properly in terms of ability; but when assignments are given, Michael is often heard to complain, "Not that again!" He will also lay his head on the table although he does get enough sleep. It is not uncommon to find "doodling" on his papers and other signs of boredom. | 50. Give him specific information about how his school work is related to life outside of school. | AI |
| | 51. Give him special assignments designed to capture his interest. | M |
| XIV. Jane has marked her table with crayons several times. Each time she was reprimanded and asked to erase the marks. Today she has put pencil marks in a library book. | 53. Try to explore her feelings about the school situation. | C |
| | 54. Take away all library privileges from Jane and ask her to erase the marks in her book as well as other library books. | D |
| | 56. Tell Jane's parents about her behavior—leave it up to them to take whatever action they feel is appropriate. | R |
| XXIII. The class is very excited about a proposed field trip. Jackie listens quietly to the hubbub and then says, "I hope it will not rain that day." At a surprise birthday party given her by the class, she whispers to Nancy, "This is so nice. I am afraid to enjoy it too much. Besides, it will all be over so soon." She often appears sad. | 89. Urge her to look at the bright side of life. | D |
| | 90. Arrange for her to participate more frequently in those activities she especially enjoys. | A |
| | 91. Get better acquainted with Jackie so as to be able to help her to discover and plan more satisfying experiences. | C |
| | 92. Ask the school counselor to accept the responsibility for this case. | R |

TABLE 2

MEANS, STANDARD DEVIATIONS, AND RELIABILITY ESTIMATES FOR THE REVISED SCALES OF TPQ, FORM 2

| Scale | Number of items per scale | *M* | *SD* | Split-half[a] reliabilities projected to 20 items |
|---|---|---|---|---|
| Disciplinarian | 18 | 68.9 | 14.69 | .86 |
| Counselor | 19 | 37.0 | 12.89 | .93 |
| Motivator | 10 | 22.6 | 6.31 | .80 |
| Referrer | 10 | 35.4 | 7.37 | .86 |
| Advice-information giver | 15 | 37.0 | 8.38 | .77 |

[a] Spearman-Brown prophecy formula.

but give some indication of the interrelationships of the scales. The reliabilities, since they are computed for data which were used in part to select the scales, must be interpreted with extreme caution. They can, however, be treated as upper bounds for the reliabilities. The reliability estimates generated for the original role dimensions from the data from Study 1 suggest that these upper bound estimates for Study 2 are not too biased. Evidence for this last point includes the fact that most of the items on the revised scales are scored on the dimensions predicted for them, suggesting that the actual reliabilities for the revised scales are probably not too different from the reliability estimates in Study 1.

TABLE 3

INTERCORRELATIONS FOR THE REVISED SCALES OF TPQ, FORM 2

| Scale | Counselor | Motivator | Referrer | Advice-information giver |
|---|---|---|---|---|
| Disciplinarian | .16 | .21 | .50 | .42 |
| Counselor | | .40 | .06 | .19 |
| Motivator | | | .33 | .38 |
| Referrer | | | | .52 |

## DISCUSSION

The results of the study indicate that it is feasible to conceptualize and measure teacher role expectations in the manner described above. The split-half reliabilities of the scales are deemed adequate. The results of the two item analyses, while resulting in some modification of the original dimensions, tend to support the concept that role dimensions "exist." Subjects who were selected at random from each of the subsamples consistently reported that the problem situations and the alternatives presented were realistic and meaningful to them.

The five revised scales have low to moderate intercorrelations. In selecting the items for these scales an attempt was made to obtain items as pure as possible, but we did not deliberately attempt to develop unrelated scales. We do not believe that the dimensions we are studying are necessarily unrelated in the behavior of people and we do not feel that it is necessary to force independence in the scales. In fact, the existence and nature of the relationships are themselves an area of study.

However, the fact of the relationships leads to another point. Since the scales are related, it is possible that fewer than five clear scales would be sufficient to treat the variance produced on the instrument. Our data and experience in working with the instrument do not point in that direction. They do lead, very definitely, to the conclusion that at this time the five revised scales as presented in this paper provide a reasonable framework for treating teacher role expectations.

Of course, the theory underlying the test and the test itself both require further investigation, but the results reported here are interpreted as indicating that the approaches which were employed are promising and that further research is warranted. At the present time, studies concerned with the following matters are in progress or under consideration: Continued work on refinement and clarification of the scales is necessary, including examination of the relationships between dimensions. More needs to be known about the determinants of teacher role expectations, the stability of teacher role expectations, and the conditions under which changes occur. We are planning to investigate the relationships between teacher role expectations and other personality variables. We want to know more about the relationship between an individual's teacher role expectations and the kind of judgments he will make of a teacher with known characteristics. Perhaps most important, we want to know more about the relationship between a teacher's concept of the teaching role and his performance, as it is observed by others.

A set of procedures which will permit the prediction of teacher effectiveness, not in terms of a set of absolute characteristics, but in terms of the traits regarded as significant by judges of differing educational philosophies, differing personal preferences, and in differing situations would have obvious value. It seems possible that this study constitutes a significant step in that direction.

## REFERENCES

RYANS, D. G. *Characteristics of teachers.* Washington, D. C.: American Council on Education, 1960.

SARBIN, T. R. Role theory. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. I. Cambridge: Addison-Wesley, 1954. Pp. 223–258.

SEARS, PAULINE S. Problems in the investigation of achievement and self-esteem motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957.* Lincoln: Univer. Nebraska Press, 1957. Pp. 265–339.

(Received May 11, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 6, 295–299

# TRAINING FIRST-GRADE CHILDREN IN CREATIVE THINKING UNDER QUANTITATIVE AND QUALITATIVE MOTIVATION[1]

CONNIE J. CARTLEDGE AND EDWIN L. KRAUSER
*Emory University*

In an experiment designed to stimulate creativity in 1st-grade children, 120 Ss who scored the lowest on a screening test of creativity were divided into experimental and control groups. Experimental Ss had 5 20-min. training sessions in how to improve toys by using Osborn's principles for developing new ideas. ½ of the experimental and control groups were motivated for number of suggestions, the other half for quality. The hypothesis of difference between trained and nontrained Ss was supported ($p < .01$) but not the hypothesis of difference between Ss motivated for quantity and quality of response.

The present experimental study is concerned with the feasibility of encouraging creative thinking in first-grade children. It is designed to test the findings of E. P. Torrance's (1961) study of creativity in elementary school children. In his experiment Torrance manipulated the following variables: (*a*) training in principles for developing creative ideas, (*b*) motivation for quantity and quality of creativity, and (*c*) competition among subjects. He concluded his study with the following statement:

> Except in the first grade, children who have had the benefit of training produced more ideas and showed more flexibility and more cleverness in their thinking than the children who had no training. . . . The results provide no support for motivating pupils in the primary grades to produce a quantity of ideas without consideration for quality [p. 41].

The problems dealt with in the present experiment arise from these concluding statements, i.e., for first-grade children, training had no effect, and differences due to motivation for quantity versus quality were inconsistent. Therefore, three null hypotheses were formulated.

Hypothesis 1:

$$M_{\text{Experimental}} - M_{\text{Control}} = 0$$

There will be no reliable difference between the scores of the children in an experimental group on a test following a 5-day training period and those of an untrained control group.

Hypothesis 2:

$$M_{\text{O}}(\text{quantity}) - M_{\text{O}}(\text{quality}) = 0$$

Hypothesis 3:

$$M_{\text{F}}(\text{quantity}) - M_{\text{F}}(\text{quality}) = 0$$

There will be no reliable difference in the Originality (O) or Flexibility (F) scores of Group A and Group B if, during training, all responses of Group A subjects are reinforced, and only the novel and original responses of Group B subjects are reinforced.

## METHOD

### *Subjects*

The subjects were 120 first-grade children selected on the basis of their low creativity scores from a larger sample of 187 children in two elementary schools.

---

[1] Based on a thesis submitted by Connie J. Cartledge (1962) to the faculty of the Graduate School of Emory University in partial fulfillment of the requirements for the degree of Master of Arts.

### Procedure

The procedure consisted of five principal steps: (*a*) administration and scoring of the initial screening test, (*b*) division into experimental groups, (*c*) administration and scoring of the initial pretests, (*d*) training under quantitative and qualitative motivation, and (*e*) administration and scoring of the individual posttests.

*The Initial Screening Test.* A nonverbal group test of creativity developed by Torrance and the staff of the University of Minnesota's Bureau of Educational Research was used in the present study. The first two tasks of Torrance's Non-Verbal Creative Thinking Tasks, Form NVA, were used as a composite to form the initial screening test. These tasks, entitled Picture Completion and Figure Completion, were administered and scored according to the instructions in the revised test manual (Yamamoto, 1962). The interjudge reliability coefficients computed from the first 50 tests for the scores Elaboration, Activity, and Flexibility range from .69 to .84. This was done by computing $\chi^2$s and converting them into contingency coefficients.

*Division into Experimental Groups.* The 128 subjects from two schools who scored the lowest on the three score distributions were assigned to the experimental and control groups such that each group was representative of the total sample. Before the next division of groups, 11 subjects dropped out and 3 were replaced. Then, the 60 left in the experimental group and the 60 left in the control group were further divided so that one half of the experimental group was trained under quantitative motivation (Group A), the other half under qualitative motivation (Group B); one half of the untrained control group was tested under quantitative motivation (Group C), and the other half of the controls was tested under qualitative motivation (Group D). One half of each of the four groups participating in the experiment was from School X; the other half was from School Y.

*The Individual Pretest.* Each subject was given Task V, the Product Improvement Test (PIT), from the verbal "How Good Is Your Imagination Test Series," Revised Form C (Torrance & Michie, 1959). The three scores which this test yields, and whose Pearson $r$ interjudge reliability coefficients range from .78 to .93, are Ideational Fluency, Flexibility, and Originality. Creativity was again defined as a composite of these scores and, except for minor changes, the scores were obtained according to the scoring manual. Fluency was the total number of responses given; Flexibility was the number of "runs" or changes in ideas in terms of the Osborn (1957) principles; and Originality was the sum of the scores weighted for their inverse frequency within the present sample. The criteria for these Originality score weights were: 1, if frequency $> 6\%$; 2, if $6\% >$ frequency $\geq 4\%$; and 3, if frequency $< 4\%$.

The testing procedure was the same for each subject, with the one exception that Groups A and C were given different instructions than Groups B and D. The subjects in Groups A and C, motivated for quantity of response, were given the following instructions:

> I would like for you to see how many ideas you can think of to improve this little toy dog so boys and girls will have more fun playing with it. I'll try to write them down just as fast as you tell them to me. Tell me every idea you can think of, it doesn't matter how good it is or how crazy it sounds. Don't worry about how much it would cost.

Subjects in Groups B and D, motivated for quality of response, were given the following instructions:

> I would like for you to see what interesting, clever, and unusual ideas you can think of to improve this little toy dog so boys and girls will have more fun playing with it. I'll try to write them down just as fast as you tell them to me. Think of as exciting ideas as you can. Don't worry about how much it would cost.

*Training under Quantitative and Qualitative Motivation.* Groups A and B were given five 20-minute training sessions which consisted of the following procedures. The experimenter put a list of the subjects' names and a picture of the toy to be used on the blackboard. The subjects' tasks were to make suggestions of how to improve the toy which was passed around so that each child could manipulate and examine it. The experimenter accepted, modified the drawing, and put a check by the subject's name on the blackboard for all suggestions from Group A. No check marks were given to the subjects in Group B whose suggestions were not accepted, and the child was told why, i.e., his responses were not very unusual or different. On the chalk ledge under the blackboard the experimenter placed a chart of the subjects' names with 5 empty squares corresponding to the 5 days of the training week to the right of the names. The subject in each group who received the most check marks pasted a blue star by his name on the group chart at the end of the training session. The experimenter told the children at the beginning of the week that the one who had the most blue stars at the end of the "contest" would receive a prize.

The experimenter taught the subjects to improve the toys by means of the Osborn principles. The following principles were presented on Monday: magnification, addition, subtraction, and multiplication. The next five were presented on Tuesday and they were: division, substitution, combination, change color, and change position. During the Wednesday session the remaining principles were introduced: reversal, sensory appeal (motion, sound, light, and odor), change shape, and change the quality of the material. Thursday's and Friday's sessions were spent in reviewing all the principles and using them to modify the toys.

Large square cards were used to teach the subjects the Osborn principles. However, in addition to the 12 card principles used by Torrance (1961), five other cards were introduced. These were: a round card to represent the principle of changing shape; an aluminum "card" to represent changing the material; and a picture of a light bulb, bell, and flower to represent changing the toy by giving it eye, ear, and smell appeal, respectively. A plain square was shown first, then each of the other cards was held up as the experimenter asked the children "What have I done to the white square now?" The principle that the card represented was then related to the suggestions the subjects had given previously and they were urged to use the principles in making further suggestions about the toy in order to win the check marks. The toys used in the sessions were the following, respectively: a toy car, a mechanical cat, a toy Nurse's or Doctor's kit (depending on what the children called it), a stuffed monkey, and a wire puzzle.

*The Individual Posttest.* After training, PIT was again administered to each subject individually.

### *Treatment of the Data for PIT*

Raw scores were transformed to difference scores by subtracting the pretest from the posttest, and these difference scores were used in all statistical analyses. Three $2 \times 2 \times 2$ mixed model factorial designs (Edwards, 1960) were computed to analyze the data of the three scores: Fluency, Flexibility, and Originality.

## RESULTS AND DISCUSSION

The means and standard error of the means of the difference scores on PIT are shown in Table 1, and the results of the analysis of variance for the training, motivation, and schools factors are shown in Table 2.

In order to increase the power of the $F$ tests some of the interaction variances have been pooled with the within-group variance according to a multistage decision rule (Winer, 1962). When pooling, care must be taken to avoid a Type II error, that is, accepting a hypothesis of zero interaction when it should be rejected. Accordingly alpha was set at the .25 level when the preliminary tests on the interaction mean squares were made. The results of these preliminary tests indicated that the mean squares of only the Motivation × Schools interaction on the tests for Fluency and Originality are significant at this alpha level. Hence the Motivation × Schools sum of squares was not pooled with the Training × Schools, Training × Motivation × Schools, and within-group sum of squares for Fluency and Originality. On the test for Flexibility, however, the Motivation × Schools interaction was pooled with these others. The Training × Motivation interaction variances were not pooled because training and motivation are fixed factors and, as such, cannot form the denominators of the $F$ ratios.

Hypothesis 1 was tested by the training factor mean square which corresponded to the effects of training versus no training averaged over the two levels of motivation and of schools. Table 2 shows that using pooled variances as denominators all three $F$ ratios are significant ($p < .01$). Training results in a superior average creativity score compared with no training for first-grade children.

The motivation mean square corre-

TABLE 1

MEANS AND STANDARD DEVIATIONS OF PIT DIFFERENCE SCORES (PRETRAINING AND POSTTRAINING) FOR THE EIGHT GROUPS STUDIED

| Test | Trained groups motivation | | | | | | | | Nontrained groups motivation | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Quantitative | | | | Qualitative | | | | Quantitative | | | | Qualitative | | | |
| | X | | Y | | X | | Y | | X | | Y | | X | | Y | |
| | $M$ | $SE_{\bar{X}}$ | $M$ | $SE_{\bar{X}}$ | $M$ | $SE_{\bar{X}}$ | $M$ | $SE_{\bar{X}}$ | $M$ | $SE_{\bar{X}}$ | $M$ | $SE_{\bar{X}}$ | $M$ | $SE_{\bar{X}}$ | $M$ | $SE_{\bar{X}}$ |
| Fluency | 6.73 | 1.34 | 2.53 | 1.86 | 7.33 | 2.42 | 5.60 | 1.57 | 5.14 | 1.32 | −1.00 | 1.38 | 2.13 | 1.47 | 1.13 | 1.20 |
| Flexibility | 4.13 | .95 | 2.73 | 1.44 | 4.40 | 1.45 | 3.20 | 1.38 | 2.67 | .81 | −.93 | 1.09 | 1.33 | .76 | −.40 | 1.01 |
| Originality | 8.67 | 1.57 | 3.27 | 2.74 | 10.27 | 3.42 | 7.20 | 2.44 | 7.27 | 2.22 | −2.27 | 2.54 | 3.40 | 2.18 | 1.53 | 1.91 |

TABLE 2

FINAL STAGE ANALYSIS OF VARIANCE $F$ RATIOS[a]

| Source | Fluency | Flexi-bility | Origi-nality |
|---|---|---|---|
| Training (A) | 10.00** | 12.55** | 7.65** |
| Motivation (B) | $<1.00$ | $<1.00$ | $<1.00$ |
| Schools (C) | 7.79** | 5.67* | 7.97** |
| A × B | $<1.00$ | $<1.00$ | $<1.00$ |
| B × C | 2.64 | — | 2.02 |

[a] After pooling.
* $p < .05$.
** $p < .01$.

sponds to a comparison between quantitative and qualitative motivation averaged over the two levels of training and of schools. For Fluency and Originality motivation was tested using the Motivation × Schools mean squares of the $F$ ratios. Flexibility was tested using the error mean square ($df = 115$) as the denominator. However, the effects of the differential motivation factor were not significant in any of the three tests, and null Hypotheses 2 and 3 were not rejected (see Table 2).

The schools mean square corresponds to a comparison between the means of School X and School Y averaged over the two levels of the other factors. It can be seen from Table 2 that a significant difference exists between the schools for all three scores: for Fluency and Originality $p < .01$; for Flexibility $p < .05$. This difference may possibly be accounted for by two factors. One is the steadily declining socioeconomic level of the neighborhood in which School Y is situated. Or, the difference may be due to time of training. At School X the subjects came early in the morning; they were fresh and ready to work. At School Y the subjects always came after recess; they were still excited and took a few minutes to settle down.

None of the interaction effects were significant, indicating that: ($a$) there exists approximately the same difference between training and no training regardless of the type of motivation and (*b*) the effect of differential motivation is independent of school regardless of the significant difference between schools.

This study suggests that with sufficient training sessions even first-grade children can benefit from training in creativity, and the results suggest some further experiments. It would, however, be desirable to determine the role of particular experimenters in producing such gains, as well as the relative permanence of the changes produced.

## REFERENCES

EDWARDS, A. L. *Experimental design in psychological research.* New York: Holt, Rinehart, & Winston, 1960.

OSBORN, A. F. *Applied imagination.* New York: Scribners, 1957.

TORRANCE, E. P. Priming creative thinking in the primary grades. *Elem. Sch. J.*, 1961, 62, 34–41.

TORRANCE, E. P., & MICHIE, H. W. *Explorations in creative thinking in the early school years: I. Scoring manual for a test of creative thinking.* (Res. memorandum BER-59-1) Minneapolis: Bureau of Educational Research, University of Minnesota, 1959.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

YAMAMOTO, KAORU. *Revised scoring manual for tests of creative thinking (Forms VA and NVA).* Minneapolis: Bureau of Educational Research, University of Minnesota, 1962.

(Received November 26, 1962)

*Journal of Educational Psychology*
1963, Vol. 54, No. 6, 300–304

# TRANSFER OF ORIGINALITY TRAINING

RICHARD C. ANDERSON
*University of Illinois*

AND

RICHARD M. ANDERSON
*Bismarck, North Dakota*

In a study of transfer of originality training, 6th-grade boys were trained to state unusual uses for a series of familiar objects in "brain storming groups." Ss who received training gave significantly more novel responses to familiar objects not included in training than control Ss. However, the training groups failed to give more solutions to insight problems involving familiar objects than the control group.

Guilford (1950) and co-workers (Wilson, Guilford, & Christensen, 1953) have suggested that frequency of uncommon responses is an indication of originality and that originality thus interpreted may be related to ability to solve "insight" problems. Maltzman, Brooks, Bogartz, and Summers (1958) presented subjects with lists of uncommon uses for objects later presented with the two-string problem. They found that exposure to uncommon responses facilitated performance on the two-string problem for women, however, the effect of training was not significant for men. Flavell, Cooper, and Loiselle (1958) found that groups that were shown one, two, three, four, or five novel uses for a microswitch showed progressively superior performance on a version of the two-string problem, in which the microswitch had to be used as a pendulum bob.

Neither Maltzman et al. nor Flavell et al. presented training to improve subjects' originality. Indeed, in each study, care was taken that none of the uncommon responses to which subjects were exposed was relevant to solving the insight problem. In neither study could the treatment be regarded as increasing the repertoire of relevant uncommon responses but rather as providing cues that elicited responses low in the existing response hierarchy.

In a pilot study (Anderson, 1961) the effects of increasing the subject's repertoire of uncommon responses were explored. Subjects were trained to state unusual uses for a series of familiar objects during "brain-storming" sessions. The brain-storming group ($N = 8$) showed significantly more ($p < .01$) originality of response to four familiar objects not included in training than a no-treatment control group ($N = 9$). The training group also showed a significantly greater ($p < .01$) number of solutions to an insight problem, invented by the authors, in which familiar objects had to be used in unusual ways, but failed to show more solutions to the two-string problem than the control group.

The study described in this paper examined further the effects of increasing the subject's repertoire of uncommon responses. Subjects were trained in brain-storming sessions to state unusual uses for each of a series of familiar objects. As a result it was expected, first, that subjects would give more uncommon responses to objects not used during training than control subjects and, second, that subjects who received training would show improved performance on insight problems involving unusual uses of familiar objects.

The experiment also explored the relationship between amount of variety in training and transfer. Previous research has shown that the larger the number of different examples of each concept encountered during training (Callantine & Warren, 1955), the greater the number of shifts from concept to concept during training (Semne & Warren, 1955), or the greater the number of tasks employed during training (Duncan, 1958), the greater the transfer of training. Amount of variety in training is likely to be an especially important variable in studies, like the present one, that are concerned with generalized rather than specific transfer of training.

## PROCEDURE

Thirty-six sixth-grade boys were randomly assigned to one of three groups. Analysis of variance showed that the IQ means, reading achievement means, and arithmetic achievement means of the three groups were not significantly different.

Two groups were trained. The control group received no treatment. Experimental subjects met in groups of three with the experimenter for 10 30-minute training sessions. At each session the subjects were shown three different objects, one at a time. Instructions were to "name as many original uses as you can" for each object. The experimenter played two roles during the sessions. First he directed the subject's attention to the physical attributes of each object. The technique was to allow the subjects to give as many responses as they would spontaneously, then summarize the attributes involved in the responses given, asking the subjects if there were any other characteristics of the object that suggested uses. The experimenter was prepared with a list of attributes for each object—such as color, texture, malleability, solubility, conductivity, weight, and length—upon which he noted responses made. When the subjects stopped responding freely, the experimenter might say, to illustrate,

> So far the uses you have named involve length, strength, and hardness. Are there any other things about this —— that suggest uses to you?

If cues of the sort just illustrated failed to elicit enough additional responses, more directive cues were given (e.g., "This —— weighs something, doesn't it? Does its weight suggest any uses to you?"). The experimenter's second role was to give verbal reinforcement (e.g., "Excellent!" "Very interesting idea!" "Good!" "Very good!") for each response he deemed original. In brief, the procedure was similar to what other investigators have called brain storming.

One group received *high-variety* training, in that 30 distinctly different objects were employed. The second group received *low-variety* training involving 30 objects falling into six different classes, with quite similar objects within classes. The six classes were fruits, boards, tin cans, lubricants, ropes, and abrasives.

After training, a Novel Uses Test was administered. Subjects were asked to list as many uses as they could for each of four objects, none of which was employed during training. It was explained that extra credit would be given for very unusual uses.

The Novel Uses Test had been given previously to 55 sixth-grade boys from an elementary school in a different community. Exhaustive and, roughly, mutually exclusive response categories were developed for each object on the basis of these data (Anderson, 1961). Each response category was given a scoring weight determined by the frequency with which it appeared. Responses that appeared in more than 15% of the protocols received no score. Responses that appeared in 11–15% of the protocols received a score of 1. Responses that appeared in 6–10% of the protocols received a score of 2. Responses contained in 1–5% of the protocols received a score of 3. Responses not contained in the content analysis scheme received a score of 4. After eliminating repetitious responses, undecipherable responses, and "nonfunction" responses, the remaining responses were categorized. A subject's total originality score consisted of the sum of the scores he received for each object.

Protocols from the Novel Uses Test were scored blind by one rater. The product-moment correlation between the total originality scores assigned by two raters in the pilot study was .98.

All subjects performed on three insight problems in separate sessions on successive days following the session in which the Novel Uses Test was given. The two-string problem was presented first. A subject was provided with a mouse trap, a basketball, an unbreakable saucer, a large handful of marbles placed in the saucer, a small plastic bag, six sheets of mimeograph paper, and a pencil. The subject was given 4 minutes to solve the problem. If he succeeded the object critical to solving the problem was taken away, or, depending upon a prearranged plan of administration (Anderson, 1961), the subject was told to solve the problem another way. The subject was given additional 3-minute periods to solve the problem in other ways until he failed to solve the problem in some 3-minute period.

The second problem, called the garbage can problem, was created by the investigators. A grate such as is found in an oven was secured to the top of a large, metal garbage can. A small metal can opener with a loop handle was placed in the bottom of the garbage can. The problem was to get the can opener out of the garbage can without tipping the garbage can or removing the grate. The subject was also furnished with a box of paper clips, a small plastic bag, two sheets of mimeograph paper, a large pair of scissors, a brick, a book, a broom handle with ends slightly rounded, a lump of modeling clay, and four 1-foot sections of glass tubing. If the subject solved the problem in the first 4 minutes he was given additional 3-minute periods to solve the problem in other ways until he failed to solve the problem in some 3-minute period. After each solution either one of the objects was taken away or the subject was told to "get the can opener out another way" depending upon a prearranged plan of administration (Anderson, 1961).

The third problem was an adaptation of Duncker's (1945) box problem that the investigators have named the candle problem. A fiberboard panel, with a piece of wrapping paper tacked to the front, was placed against a wall. The following objects were scattered on a nearby table: a paper cup full of wooden matches, a paper cup full of thumb tacks, a paper cup full of small candles, a screw driver, a synthetic sponge, a wooden ruler, a scissors, a paper punch, and a box of paper clips. When the subject was brought into the room, the experimenter made three marks on the sheet of paper attached to the front of the fiberboard panel and then asked the subject to "fix a candle on each of the spots I just marked" in such a way that the candles could burn without scorching the paper. Five minutes were allowed to solve the problem. If successful, the subject was allowed additional 4-minute periods until he failed to solve the problem in some 4-minute period. After any solution, either one of the objects was removed or the subject was asked to solve the problem another way, depending upon a prearranged plan of administration (Anderson, 1961).

The insight problems were presented by an experimenter who did not know which subjects were members of a training group and which were members of the control group. None of the objects involved in the insight problems was used during training. The problems were presented to all subjects in the same order, namely, the two-string problem, the garbage can problem, and the candle problem. It was necessary to use the same order of presentation, instead of a balanced-random order, so that the subjects who had performed on problems could be prevented from contaminating those who had not.

## RESULTS

Differences among groups (see Table 1) in scores on the Novel Uses Test were highly significant ($F = 33.69$, $n_1 = 2$, $n_2 = 33$, $p < .001$). The mean originality of each training group was significantly higher ($p < .001$) than the mean originality of the control group. The difference in originality of the high-variety and low-variety groups was not significant.

The number of solutions to each of the problems showed Poisson or near-Poisson distributions. After a square root transformation had been applied, Hartley's test indicated that there were not significant differences among variances and chi square goodness-of-fit tests showed that the distributions did not vary significantly from normality. The mean and standard deviation of number of solutions (transformed) to each of the problems for each of the groups appear in Table 1.

Analysis of variance (Lindquist, 1953, Type I) indicated that there were significant differences among problems. The $F$ ratios for the among-groups effect and the interaction were both less than 1.00.

Total-sample correlations based on transformed scores of .20, .26, and .13 appeared between the two-string problem and the garbage can problem, the two-string problem and the candle prob-

TABLE 1

MEANS AND STANDARD DEVIATIONS OF PROBLEM SOLVING PERFORMANCE AND ORIGINALITY SCORES

| Group | Number of solutions [a] | | | | Originality score |
|---|---|---|---|---|---|
| | Two-string problem | Garbage can problem | Candle problem | All problems | |
| High variety[b] | | | | | |
| *M* | 1.31 | 1.37 | .76 | 3.45 | 98.0 |
| *SD* | .91 | .91 | .89 | 1.87 | 23.7 |
| Low variety[b] | | | | | |
| *M* | 1.38 | 1.24 | .91 | 3.52 | 89.4 |
| *SD* | .64 | .48 | .95 | 1.41 | 29.9 |
| Control[b] | | | | | |
| *M* | 1.11 | 1.21 | .67 | 2.99 | 22.1 |
| *SD* | .79 | .57 | .87 | 1.31 | 19.8 |
| All groups | | | | | |
| *M* | 1.27 | 1.27 | .78 | | |
| *SD* | .74 | .64 | .85 | | |

[a] Square root of number of solutions.
[b] $N = 12$.

lem, and the garbage can problem and the candle problem, respectively.

## DISCUSSION

The training procedures were notably successful in increasing the number of novel uses subjects listed for familiar objects. The result is similar to that obtained by Maltzman, Bogartz, and Berger (1958), who, having trained subjects to make uncommon word associations, found a significant increase in the number of uncommon associations to test words. Both the Maltzman, Bogartz, and Berger study and the present one suggest that transfer among highly similar tasks can be achieved in the area of "creativity" or "imagination."

TABLE 2

ANALYSIS OF VARIANCE OF PROBLEM SOLVING PERFORMANCE

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Groups (G) | 2 | .33 | < 1.00 |
| Error | 33 | .76 | |
| Total between subjects | 35 | | |
| Problems (P) | 2 | 2.89 | 5.25* |
| G × P | 4 | .09 | < 1.00 |
| Error | 66 | .55 | |
| Total within subjects | 72 | | |

* $p < .01$.

Training did not transfer to the insight problems. Indeed, the *F* ratios for both the among-groups effect and the groups × problems effect were less than one. Though the training groups did learn to produce unusual responses, the training was unsuccessful in increasing the likelihood that the subjects would give the particular unusual responses required to solve the problems.

Correlations of .01 and .70 appeared between originality and the square root of number of solutions to all problems for the pooled training groups and the control group, respectively. Though the latter coefficient was not significantly different from zero because of the small sample, a similar pattern of relationships was observed in the pilot study where the correlation for the training group was .05 and the correlation for the control group was .31. Note that the Novel

Uses Test was administered after training in each case. Apparently the training interferred with the relationship between originality and performance on insight problems.

The low correlations among the insight problems indicate either that the problems are very unreliable measures or that something quite different is required for the solution of each of them, both factors that would make a demonstration of transfer difficult. And, of course, the sample was small and it may be that more intensive training or training of a different sort would result in transfer. Still, the present study does not support the optimism, prevalent in some circles, that "massive generalized transfer of training" can be achieved in the areas of "intuition" and "imagination." At least, this study suggests that such transfer is going to be very difficult to demonstrate, requiring large samples, more reliable measures than have traditionally been used in the area, and rather intensive, carefully conceived training procedures.

## REFERENCES

ANDERSON, R. M. Positive transfer of an imagination set. Unpublished master's thesis, Rutgers University, 1961.

CALLANTINE, M. F., & WARREN, J. M. Learning sets in human concept formation. *Psychol. Rep.*, 1955, **1**, 363–367.

DUNCAN, C. P. Transfer after training with single versus multiple tasks. *J. exp. Psychol.*, 1958, **55**, 63–72.

DUNCKER, K. On problem-solving. *Psychol. Monogr.*, 1945, **58**(5, Whole No. 270).

FLAVELL, J. H., COOPER, A., & LOISELLE, R. H. Effect of number of preutilization functions on functional fixedness in problem-solving. *Psychol. Rep.*, 1958, **4**, 343–350.

GUILFORD, J. P. Creativity. *Amer. Psychologist*, 1950, **5**, 444–454.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

MALTZMAN, I., BOGARTZ, W., & BERGER, L. A procedure for increasing word association originality and its transfer effects. *J. exp. Psychol.*, 1958, **56**, 393–398.

MALTZMAN, I., BROOKS, L. O., BOGARTZ, W., & SUMMERS, S. S. The facilitation of problem solution by prior exposure to uncommon responses. *J. exp. Psychol.*, 1958, **56**, 399–406.

SEMNE, L. R. M., & WARREN, J. M. Proactive facilitation of performance on the Wisconsin Card Sorting Test. *Psychol. Rep.*, 1955, **1**, 433–436.

WILSON, R. C., GUILFORD, J. P., & CHRISTENSEN, P. R. The measurement of individual differences in originality. *Psychol. Bull.*, 1953, **50**, 362–370.

(Received January 1, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 6, 305–315

# CHANGES IN CRITICAL THINKING, ATTITUDES, AND VALUES FROM FRESHMAN TO SENIOR YEARS[1]

IRVIN J. LEHMANN
*Michigan State University*

To investigate changes in critical thinking ability, stereotypic beliefs, dogmatism, and values, a battery of cognitive and affective measures was administered to 1051 students as freshmen and then as seniors. The data were considered separately for males and females. The principal findings were that: (a) there is a significant decrease in stereotypic beliefs and unreceptivity to new ideas, (b) seniors are more "outer-directed" than they were as freshmen, (c) there is a significant improvement in critical thinking ability, and (d) seniors are not more homogeneous in certain attitudinal traits than freshmen. It was also shown that most of the changes took place during the freshman and sophomore yrs.

During the past decade, considerable interest has been generated among psychologists, sociologists, and educators in the study of attitudes and values. Since the advent of Newcomb's (1943) work in the 1930s, major longitudinal studies of the personality characteristics of college students have been and are being conducted at Vassar College (Sanford, 1956), Sarah Lawrence College (Murphy & Raushenbusch, 1960), Cornell University (Goldsen, Rosenberg, Williams, & Suchman, 1960), Michigan State University (Lehmann & Ikenberry, 1959), San Jose State College (Plant, 1962), and Santa Barbara College (Foster, Stanek, & Krassowski, 1961), to name a few. Although the methodology and instrumentation employed might have differed from one institution to another, they all shared the common thesis of studying the personality development of college students.

Although it is sometimes asserted that attitudes and values are instilled early in life and are most easily modifiable in infancy and adolescence, curriculum planning in our colleges and universities assumes that attitudes and values are still modifiable at age 17–22 or older. In fact, if we define education in its broadest sense, we might expect that there would be a change or reinforcement of certain ideals, beliefs, and interests of students from their freshman to senior years.

Contradictory evidence has been presented regarding the impact of college on student attitudes and values (Jacob, 1958; Newcomb, 1943; Webster, 1958). At the same time, data have indicated that from the freshman to senior years, changes in attitudes and values will occur (Newcomb, 1943; Webster, 1958), but that the degree and extent to which attitudes and values are modifiable depend upon the nature of the experience (Smith, 1955), the type and nature of the contact (McGuigan, 1958), the personality make-up of the individual (Nelson, Blake, Morton, & Olmstead, 1956), the group's approval of new attitudes (Rosenberg, 1956), and the subject's perception of the outcome (Carlson, 1956).

---

[1] The research reported in this article was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare. A more detailed report is to be found in the final report *Critical Thinking, Attitudes, and Values in Higher Education* (Lehmann & Dressel, 1962).

One of the major faults of many so-called longitudinal studies has been a lack of any periodic study of the original population between the initial and final stages. Although this study does not claim to have considered or circumvented all of the problems involved in longitudinal studies (for example, absence of a noncollege control group), an attempt has been made to follow up the original population or random samples of it at various stages of their college careers.

The purpose of this paper is not to demonstrate whether college aids or facilitates the development of attitudes and values, nor are we concerned at this time with identifying those factors which might have an impact upon student personality development. We are primarily concerned with studying what changes, if any, occur in college students' critical thinking ability; attitudes of stereotypy and dogmatism; traditional value orientation; and religious, social, and political views from their freshman to senior years.

## DESCRIPTION OF INSTRUMENTS USED

One instrument used was the Inventory of Beliefs (American Council on Education, 1950) which explores the students' tendency toward (*a*) ideocentrism, (*b*) ethnocentrism, (*c*) sociocentrism, and (*d*) egocentrism. It consists of 120 clichés, such as "a person has troubles of his own; he can't afford to worry about other people," "the best government is one which governs least," "the worst danger to real Americanism during the last 50 years has come from foreign ideas and agitators." The subjects respond to each item by means of a 4-element key: strongly agree, agree, strongly disagree, disagree.

A high scorer is thought to be mature, flexible, adaptive, and democratic in his relationships with others; a low scorer immature, rigid in outlook, compulsive, and authoritarian in his relationships with others. Implicit in the development of this scale and the choice of items was the assumption that general education makes individuals less stereotypic in their beliefs. Some experience in general education courses should therefore raise the scores of students.

A second instrument used was the Test of Critical Thinking (American Council on Education, 1953) which consists of 52 verbal, situational-type problems designed to measure five factors thought to be involved in critical thinking: the ability to (*a*) define a problem, (*b*) select pertinent information, (*c*) recognize stated and unstated assumptions, (*d*) formulate and select relevant hypotheses, and (*e*) draw valid conclusions. It would be expected that an effective general education course would cause students to raise their scores.

The Differential Values Inventory (Prince, 1957) which places a student's values along a traditional-emergent continuum was chosen as a third instrument. The scale consists of 64 pairs of statements, each pair containing a traditional- and emergent-value oriented item.[2] The traditional-value orientation or the emergent-value orientation of a student is determined by his choices from among 64 pairs of statements such as "Attain a higher position than my father or mother" (traditional) versus "Enjoy more of the good things in life than my father or mother enjoyed" (emergent value); "Feel that children are born good" (emergent) versus "Feel that children are born sinful" (traditional value). Scores can range from 0 to 64 for the total test and from 0 to

---

[2] Not all the pairs of statements contain polar items. For example, a future time orientation item (traditional) may be pitted against a sociability item (emergent) even though the polar item is a present time orientation.

16 for each of the subtests. A high score indicates a leaning toward traditional values; that is, the individual places high value on personal respectability, respect for others, feelings of guilt, thrift, and self-denial. A high scorer values hard work as good in itself and necessary for success, places his personal and individual desires and beliefs equal to or above the desires and ideas of the group, and is willing to sacrifice present needs for future reward and satisfaction. A low score indicates a leaning toward the emergent values, that is, toward the importance of getting along with others and achieving group harmony. A low scorer questions all absolutes, accepts group determined morality, allows his actions to be governed primarily by the consideration of others, and sacrifices future goals for present needs. Using Riesman's classification, one might call the traditional values "inner directed" and the emergent values "outer or other directed." The scale yields a total score and eight subscores: values of Puritan morality, individualism, work-success ethic, and future time orientation (the traditional-value score); sociability, conformity, relativism, and a present time orientation (the emergent-value score).

The Dogmatism Scale (Rokeach, 1960) was used to measure students' receptivity to new ideas, their degree of open mindedness, and their authoritarian attitude. This scale differs from the traditional California E and F Scales in that it does not attempt to tie in authoritarianism with conservatism or ethnocentrism. The Dogmatism Scale consists of 40 items such as "To compromise our political opponents is dangerous because it usually leads to the betrayal of our own side," or "It is only natural for a person to be fearful of the future," to which subjects respond by means of a 6-element "agree-disagree" key (there is no neutral position). High scorers are thought to be dogmatic, closed minded, and unreceptive to new ideas.

An Experience Inventory[3] especially constructed for this study (consisting of 103 items) was designed to measure social and political attitudes such as "The United States should continue nuclear testing in the atmosphere," "Medical care for the aged should be provided by the Federal government," as well as attitudes towards the objectives of college (vocational versus liberal arts orientation), factors important for gaining prestige with students and faculty, and so forth. For the purpose of this study, we will consider only these items that relate to some behavioral change (either differences measured by a pretest and posttest or those mentioned by students as self-perceived changes).

## METHOD

The instruments previously described, with the exception of the Experience Inventory, were administered during orientation week to all freshmen (1,436 males and 1,310 females) entering Michigan State University in the fall of 1958. The only restrictions placed upon the population was that they be native-born Americans and attending college for the first time. Near the end of their senior year, this same battery of cognitive and affective tests, as well as the Experience Inventory, was administered to those students remaining in order to learn what changes had taken place during their 4 years at Michigan State University. The analysis of data is restricted to those students for whom we have usable pretest and posttest data. The findings and conclusions presented in this paper are based upon a sample of 1,051 students (70% of the original group still enrolled in college near the end of their senior year[4] and 39% of the original sample).

---

[3] Copies may be obtained from the author.

[4] The test-retest sample consisted of 590 males and 461 females. Tests of significance indicated that there was no significant difference in initial status between these seniors who participated and those who did not participate in the senior year retest.

The $t$ test for correlated sample means was used to analyze changes in critical thinking, stereotypy, dogmatism, and traditional-value orientation scores for the population from their freshman to senior years. Because of initial sex differences on the cognitive and affective measures studied, the data were analyzed separately for males and females.

## FINDINGS AND DISCUSSION

The data relative to changes in critical thinking processes, attitudes of stereotypy and dogmatism, and traditional-value orientation are presented in Tables 1–3. The data concerning *self-perceived* changes in typology and in attitudes and values are presented in Tables 4 and 5.

Reference to Tables 1 and 2 reveals that for both males and females, there were significant changes in critical thinking ability, attitudes, and values from the freshman to senior years. The degree and direction of changes might be summarized as follows:

1. Both males and females received significantly higher scores on the Test of Critical Thinking at the end of 4 years of college than they did when they entered Michigan State University as freshmen. This would suggest that these students became more proficient in such tasks as the selection of pertinent information, the ability to draw generalizations and make conclusions, and the ability to formulate and select relevant hypotheses. There does not appear to be any evidence that one sex or the other changed more in this trait during their 4 years at college.

2. Both males and females became less stereotypic in their beliefs from their freshman to senior years. In other words, these students became more flexible, less rigid, and less authoritarian during their 4 years at college. It would appear that the females underwent a more marked change than the males during this period. Although an analysis of covariance was not performed, the fact that there is no significant sex difference initially, but a marked sex difference at the end of the senior year does suggest, at least, that the females changed more than the males. This is indeed difficult to explain in terms of curricular differences but might be more easily explained in terms of student subculture differences between the sexes. Females, on the whole, both at the beginning and end of college tend to be more oriented to conformity and sociability—to doing things to please others.

TABLE 1

MEANS, STANDARD DEVIATIONS, AND TESTS OF SIGNIFICANCE FOR CHANGES IN CRITICAL THINKING ABILITY, ATTITUDES, AND VALUES BETWEEN FRESHMAN AND SENIOR YEARS

| Measure | Males ($N = 590$) | | | | | Females ($N = 461$) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $M$ | | | | | $M$ | | | | |
| | Pretest | Posttest | $r$ | $t^1$ | $t^2$ | Pretest | Posttest | $r$ | $t^1$ | $t^2$ |
| Inventory of Beliefs | 64.77<br>(14.10) | 73.08<br>(15.10) | .61 | 15.48* | 1.93 | 65.99<br>(13.11) | 78.67<br>(14.91) | .67 | 23.66* | 3.72* |
| Test of Critical Thinking | 32.82<br>(6.75) | 39.72<br>(6.11) | .63 | 27.60* | 3.12* | 33.07<br>(6.88) | 39.45<br>(6.60) | .70 | 26.15* | 1.25 |
| Differential Values Inventory | 34.84<br>(7.03) | 32.38<br>(7.61) | .43 | 37.27* | 2.14* | 34.22<br>(6.87) | 29.60<br>(6.84) | .45 | 13.79* | 0.11 |
| Dogmatism Scale | 166.85<br>(25.61) | 153.98<br>(22.94) | .53 | 13.21* | 3.15* | 162.87<br>(25.14) | 146.69<br>(24.16) | .50 | 14.07* | 0.98 |

Note.—$t^1$ = test of significance for correlated means; $t^2$ = test of significance for correlated variances. *SDs* given in parentheses.
* $p < .05$.

TABLE 2

MEANS AND STANDARD DEVIATIONS FOR CRITICAL THINKING ABILITY, ATTITUDES, AND VALUES FOR MALES AND FEMALES AT DIFFERENT STAGES OF COLLEGE

| Measure | Beginning of freshman year[a] (Male $N = 590$, female $N = 461$)[c] | | End of freshman year[a] (Male $N = 590$, female $N = 461$)[c] | | End of sophomore year[b] (Male $N = 197$, female $N = 217$)[c] | | End of junior year[b] (Male $N = 235$, female $N = 189$)[c] | | End of senior year[a] (Male $N = 590$, female $N = 461$)[c] | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Inventory of Beliefs | 64.77<br>(65.99) | 14.10<br>(13.11) | 67.00<br>(69.32) | 14.68<br>(14.86) | Not given<br>(Not given) | | 73.68<br>(75.92) | 13.67<br>(15.51) | 73.08<br>(78.67) | 15.10<br>(14.91) |
| Test of Critical Thinking | 32.82<br>(33.07) | 6.75<br>(6.88) | 37.51<br>(37.48) | 6.41<br>(6.53) | 39.06<br>(38.53) | 6.04<br>(6.31) | 38.99<br>(38.96) | 6.47<br>(6.71) | 39.72<br>(39.45) | 6.11<br>(6.60) |
| Differential Values Inventory | 34.84<br>(34.22) | 7.03<br>(6.87) | 33.07<br>(32.12) | 7.53<br>(7.04) | 32.25<br>(30.27) | 7.81<br>(6.87) | 32.47<br>(29.76) | 8.18<br>(6.52) | 32.38<br>(29.60) | 7.61<br>(6.84) |
| Dogmatism Scale | 166.85<br>(162.87) | 25.61<br>(25.14) | Not given<br>(Not given) | | 156.31<br>(151.03) | 24.21<br>(24.04) | Not given<br>(Not given) | | 153.98<br>(146.69) | 22.94<br>(24.16) |

[a] Same subjects.

[b] Random sample of eligible population enrolled at time of testing used. Although there might be some slight bias inasmuch as some of the subjects could have withdrawn before their senior year, the number involved is minimal.

[c] First set of figures for males; those directly below in parentheses for females.

In addition, they appear to be more susceptible to outside factors insofar as their behavior is concerned—that is, they tend to be more outer or other directed than males. This latter thesis was supported by an analysis of those experiences college seniors felt had a very modifying or reinforcing effect on their behavior. From this analysis, it is evident that a proportionately larger percentage of females than males felt their friends, roommates, and discussions or "bull-sessions" had a very marked impact upon their attitudes, values, beliefs, and interests.

Because of their close attachment to a group, because of their initial personality disposition, and because they are more open minded and receptive to new ideas than males, it is quite conceivable that females will be influenced or affected to a greater degree than males by the mores and demands of the peer group. What should be remembered, however, is that the direction of change for both males and females is toward a lessening of stereotypic beliefs and prejudices.

3. Both males and females became less traditional-value oriented from their freshman to senior years. In other words, they became less inner directed and more outer directed. They tended to become less absolute in their moral, religious, and ethical beliefs; had less regard for the "Protestant ethic" of Puritan morality; became more questioning; and became more conforming to peer mores. Once again, it would appear that the females were more susceptible to change than the males. Possibly, the explanations advanced earlier are also true here. This does suggest that more research is needed to try and ascertain the reasons for this phenomenon.

4. Both males and females became more open minded and/or more receptive to new ideas from their freshman to senior years, the females undergoing more marked change than the males.

5. Although Jacob (1958) concluded that college students tended to become more homogeneous[5] in their values from their freshman to senior years, the re-

---

[5] Increased homogeneity would be expected due to withdrawals from college.

sults of this study as well as those conducted at Vassar (Sanford, 1956), do not completely support this view. Reference to Table 1 reveals that for the males, there was significantly greater heterogeneity in traditional values (Differential Values Inventory) and significantly more homogeneity in receptivity to new ideas (Dogmatism Scale) at the end of the senior year than at the beginning of the freshman year. For the females, there is significantly greater heterogeneity in stereotypic beliefs (Inventory of Beliefs) after 4 years of college. In other words, in only one instance, male dogmatism, was there evidence to suggest a greater homogeneity of values in the senior year than in the freshman year. It is conceivable, however, that with more reliable and general measures than those used by previous investigators of attitude change, increasing variances will be expected with age, at least to a certain age (McConnell, 1934; Matteson, 1955; Strong, 1943). In agreement with this theory, the standard deviations of the various college classes vary significantly, rising and falling for freshmen, sophomores, juniors, and seniors. This trend was also evident in the Vassar data reported by Webster (1958).

Initially, it would appear from the analysis of pretest and posttest data that significant changes in critical thinking ability, attitudes, and values occurred during the students' 4 years at Michigan State University. This, however, is not the case. For the most part, the greatest magnitude of change occurred during the freshman and sophomore years as is revealed in Table 3. For the changes in value orientation (from traditional to emergent) and critical thinking ability, it would appear that the greatest overall change took place sometime between the beginning and end of the freshman year.[6] It is conceivable, of course, that all or most of this initial change might have occurred during the first few months of the freshman year and even during the first few weeks insofar as the students' values are concerned. What is significant, however, is that looking at changes between the freshman and senior years can be very misleading.

### *Self-Perceived Changes*

From the analysis of the Experience Inventory data, it is readily evident that

TABLE 3

MEAN CHANGES IN CRITICAL THINKING ABILITY, ATTITUDES, AND VALUES FOR MALES AND FEMALES AT DIFFERENT STAGES OF COLLEGE

| Measure | Fall 1958 to spring 1959 | Fall 1958 to spring 1960 | Fall 1958 to spring 1961 | Fall 1958 to spring 1962 |
|---|---|---|---|---|
| Inventory of Beliefs | 2.33[a]<br>(3.33) | —<br>— | 8.89<br>(9.93) | 8.31<br>(12.68) |
| Test of Critical Thinking | 4.69<br>(4.41) | 6.24<br>(5.46) | 6.17<br>(5.89) | 6.90<br>(6.38) |
| Differential Values Inventory | −1.73<br>(−2.10) | −2.59<br>(−3.95) | −2.37<br>(−4.46) | −2.46<br>(−4.62) |
| Dogmatism Scale | —<br>— | −10.54<br>(−11.84) | —<br>— | −12.87<br>(−16.18) |

[a] First set of figures are males; those directly below in parentheses for females.

---

[6] It would appear that, after the junior year, a plateau was reached with respect to the value orientation of these college students.

these students perceived a marked change in their attitudes and values. In one section of the senior-year Experience Inventory, students were asked to classify themselves into one of four typologies: vocational, academic, nonconformist, and collegiate (see Footnote a to Table 4). In addition, they were asked to recall their behavior when freshmen and classify themselves into one of these types. Reference to Table 4 reveals that when these students were seniors, they felt that they were less concerned with attending college for vocational preparation than when they entered as freshmen in 1958. As seniors, they felt strongly concerned with the academic and social aspects of college life as well as the vocational. It should also be noted that the percentage of students who might be described as trying to get by with a minimum effort decreased over the 4-year period. In addition, a proportionately larger percentage of seniors (in contrast to how they stated they felt as freshmen) said that the social life in college was not the most important but was certainly essential for one's general development.

Table 5 contains the responses to a variety of items (found in the senior-year Experience Inventory) describing changes (or lack of change) which seniors felt came about from their freshman to senior years in such areas as tolerance, religion, the value of a college education, and so forth. For each of the 39 items, the student was asked "In what way are you different *now* from what you were as a freshman?" He responded to each item by means of a 3-element key: I possess *more* of this quality; I possess the *same* amount of this quality; I possess *less* of this quality. From Table 5 one might conclude that:

1. A majority of students indicated that they have changed in one direction or another in their attitudes, values, opinions, and beliefs from their fresh-

TABLE 4

SENIORS' SELF-PERCEIVED CHANGES FROM FRESHMAN TO SENIOR YEARS

| Item | Type and number responding | | | |
|---|---|---|---|---|
| | W | X | Y | Z |
| Which of the above (W, X, Y, Z)[a] comes *closest* to describing the kind of person you consider yourself to be *now*? | 320<br>(29.1%) | 215<br>(19.6%) | 437<br>(39.8%) | 123<br>(11.2%) |
| Which of the above is *least* descriptive of the kind of person you consider yourself to be *now*? | 210<br>(19.1%) | 278<br>(25.3%) | 41<br>(3.7%) | 566<br>(51.9%) |
| Which of the above comes *closest* to describing the kind of person you were when you first came to MSU? | 459<br>(41.9%) | 173<br>(16.0%) | 258<br>(23.1%) | 205<br>(19.0%) |
| Which of the above comes *closest* to describing the kind of person you *would like to be* if you had the choice? | 114<br>(10.4%) | 330<br>(30.0%) | 604<br>(54.9%) | 44<br>(4.0%) |

Note.—The table is to be read as follows: Whereas 41.9% of the seniors felt that as freshmen they fell in Category W, 29.1% of these students as seniors felt they fell in Category W. The first and third items should be read *vertically* and not horizontally.

[a] *Type*: W—For the most part, this person's primary reason for being in college is to obtain vocational or occupational training. X—For the most part, this person would consider himself to be someone who is primarily motivated by intellectual curiosity. Y—For the most part, this person feels that the social life in college is not the most important but is certainly essential for his general development. Z—For the most part, this person attempts to "make grades" but will rarely go out of his way to do extra or non-assigned readings.

TABLE 5

SELF-PERCEIVED CHANGES IN ATTITUDES, VALUES, AND VIEWS FROM FRESHMAN TO SENIOR YEARS

| In what ways are you different *now* from what you were as a freshman? | Percentage responding[a] | | |
|---|---|---|---|
| | More | Less | Same |
| Tolerance of people differing in race, creed, color, or religion | 68 (73) | 7 (5) | 27 (22) |
| Respect for the views and opinion of other people | 80 (82) | 3 (2) | 17 (16) |
| Respect for views and opinions opposite to mine | 76 (80) | 4 (4) | 20 (16) |
| Tolerance of unconventional dress, behavior, and manners | 48 (64) | 21 (13) | 31 (23) |
| Insight into the behavior of other people | 85 (93) | 2 (1) | 13 (6) |
| Ability to get along with other people | 73 (73) | 3 (3) | 24 (24) |
| Interest in political matters | 60 (60) | 7 (8) | 33 (32) |
| Interest in social issues | 68 (71) | 5 (4) | 27 (25) |
| Interest in intellectual and cultural matters | 73 (84) | 4 (2) | 23 (14) |
| Interest in scientific developments | 70 (60) | 6 (5) | 24 (35) |
| Interest in world affairs | 77 (76) | 2 (3) | 21 (21) |
| Ability to adjust to conditions not to my liking | 73 (74) | 5 (6) | 22 (20) |
| Ability to accept disappointment | 64 (67) | 5 (4) | 31 (29) |
| Ability to change my views in the presence of facts | 80 (78) | 1 (2) | 19 (20) |
| Confidence in my ability to deal with new problems | 82 (83) | 4 (4) | 14 (13) |
| Responsibility for my own behavior | 80 (80) | 1 (1) | 19 (19) |
| Dependence on class attendance for learning | 28 (28) | 43 (39) | 29 (33) |
| Feeling that the quality of one's education depends on the institution rather than the individual | 13 (10) | 63 (67) | 24 (23) |
| Dependence on my age group for behavior patterns | 11 (10) | 56 (62) | 33 (28) |
| Acceptance of the Bible as a guide to modern living | 11 (13) | 34 (31) | 55 (56) |
| Attachment to a religious sect or denomination that I can believe in and defend | 12 (17) | 37 (34) | 51 (49) |
| Respect for law | 34 (25) | 11 (7) | 55 (68) |
| Respect for rules and regulations | 28 (18) | 23 (27) | 49 (55) |
| Respect for persons in positions of authority | 35 (20) | 16 (18) | 49 (62) |
| Feeling that money is of primary importance | 12 (11) | 38 (37) | 50 (52) |
| Desire to accept a job for the satisfaction it has to offer rather than the salary it pays | 61 (58) | 9 (8) | 30 (34) |
| Drive to get ahead as quickly as possible | 50 (27) | 14 (24) | 36 (49) |
| Feeling that a major aim of college is to prepare one for a vocation or profession | 33 (20) | 36 (45) | 31 (35) |
| Feeling that a college should also stress a liberal-arts-type of education | 66 (68) | 9 (6) | 25 (26) |
| Feeling that a college education is necessary to succeed in the world | 59 (59) | 13 (14) | 28 (27) |
| Importance of grades as measures of achievement | 18 (11) | 61 (66) | 21 (23) |
| Confusion as to what I want out of life | 21 (21) | 56 (64) | 23 (15) |
| Optimistic outlook for my future | 64 (68) | 10 (10) | 26 (22) |
| Pessimistic outlook for my future | 8 (8) | 67 (71) | 25 (21) |
| Optimistic outlook for future of civilization | 45 (39) | 20 (29) | 35 (32) |
| Pessimistic outlook for the future of civilization | 20 (27) | 47 (43) | 33 (30) |
| Feeling of the necessity for religious faith for living in modern times | 33 (40) | 29 (23) | 38 (37) |
| Commitment to a set of religious beliefs | 20 (24) | 34 (33) | 46 (43) |
| Awareness of my goals in life | 75 (82) | 8 (9) | 17 (9) |

[a] First set of figures is for males ($N = 645$); figures in parentheses are for females ($N = 439$).

man to senior years. For every item, however, it is readily evident that some students believed they did *not* change in that particular respect. This, incidentally, might be desirable for lack of change in some people is as desirable as change in others.

2. For the majority of items, there were no appreciable differences in the proportion of males and females responding "I possess more, the same, or less of this quality *now* than I had as a freshman." Some of the marked discrepancies were as follows: A proportionately larger number of females than males stated that since their freshman year, they *felt* that they had become *more* interested in intellectual and cultural matters; *less* inclined to regard that the major aim of college was to prepare one for a vocation or profession; *less* confused as to what they wanted out of life; *more* aware of their goals in life; *less* dependent on their peer group for behavior patterns; *more* tolerant of unconventional dress, behavior, and manners; and *more* tolerant of people differing in race, creed, color, and religion.

3. A small percentage of students felt that they changed in what might be termed a negative direction. For example, some students indicated that they became less tolerant of people differing in race, creed, or color; that they developed less respect for law, rules, and regulations; and that they had a less optimistic outlook for their future.

4. Students became more respectful of the views and opinions of other people while they were at college. Although there had not always been a change in their views and opinions with regard to some issues such as desegregation, admittance of Red China to the United Nations, or continuation of nuclear testing in the atmosphere, it is readily evident that a sizable portion of the students developed the ability to listen to the opinions and views of others before drawing any conclusions.

5. Although a proportionately larger number of both males and females felt that as seniors (in contrast to their behavior as freshmen) they had developed greater interest in political matters, social issues, and world affairs, these data might be misleading. For example, in the interviews with sophomores and juniors, it was readily evident that these students were self-centered and rather apathetic insofar as world affairs were concerned. In fact, the only time they became interested or concerned was when the world situation threatened to involve them personally. For example, they expressed concern about the Congo and Katanga because many felt this might develop into another Korea and that they would be drafted. Other than this, they had very little interest in world affairs.

6. A proportionately larger number of both males and females in their 4 years at college was weaned from home and became independent. They developed greater confidence to deal with new problems, became more responsible for their own behavior, and less dependent on their peers for behavior patterns. No doubt similar findings would have been obtained had a noncollege population been studied. It would appear that these changes are no more than a reflection of the maturation process. It is conceivable, of course, that college attendance acted as a catalyst.

7. In reference to changes in religious values and/or the value or need for religion, it would appear that the data are contradictory since people might change in one but not in the other or vice versa. For example, after 4 years of college, 36% of the seniors stated that they became *less* attached to a religion that they can believe in and defend while

14% of the seniors felt that in this period they became *more* attached; 34% stated they became *less* committed to a set of religious beliefs versus 22% who felt they became *more* committed; nearly three times as many students (363 versus 132) became *less* inclined rather than more inclined to accept the Bible as absolute and infallible. Nevertheless, these changes are not accompanied (in direct proportion) to a decrease in the numbers who have less feeling for the need for religion. If this were the case, we should have seen about twice as many of these students feeling less (rather than more) need for religion and this is not so—35% stated that they felt more need for religion versus 26% of the sample who felt they have less need.

8. For a large percentage of these students, money was now subordinate to the satisfaction to be gained from a job. At the same time, a proportionately larger number of males than females felt that they had more drive to get ahead as quickly as possible. This finding is rather difficult to reconcile with the previous one inasmuch as one normally tries to advance rapidly for monetary gain.

9. A proportionately larger number of seniors (in contrast to their views as freshman) no longer felt that the major aim of college is to prepare one for a vocation but that each student should be provided with a liberal education.

## CONCLUSIONS

There is no denying the fact that changes in critical thinking ability, attitudes, and values occur from the student's freshman to senior years in college. It is difficult, however, in the absence of a noncollege control group to relate these changes directly to college education. In fact, in interviews with sophomores and juniors, it is evident that the informal, nonacademic experiences such as friends, persons dated, "bull-sessions," and so forth have a greater impact upon personality development than do the formal, academic experiences such as courses and instructors. Moreover, it is only after the students entered their major that any evidence of the impact of formal academic experiences appeared.

For these college students, the greatest changes in attitudes, values, and critical thinking ability take place during the freshman and sophomore years. What is significant at this time in the students' lives? Are students more susceptible to the influence of the student subculture and its mores at this time? Is dormitory living, with its interaction of different personalities, beliefs, and interests, a partial explanation? Would the same changes (degree and direction) have occurred had these students not attended college?[7]

Since the informal, nonacademic experiences are mentioned by these students as being more potent during their freshman and sophomore years than are the formal, academic experiences, it is conceivable that the observed changes are not necessarily due to educational activities deliberately undertaken by the college. Even if the maturation process is the single, most important factor at this time colleges and universities should still attempt to isolate and to provide those experiences which are most effective in influencing personality development.

---

[7] Plant (1962) reported no significant decrease in ethnocentrism and dogmatism after 4 years between a college and noncollege population. Brown and Bystryn (1956) found that time spent on college had little bearing on authoritarian changes.

## REFERENCES

AMERICAN COUNCIL ON EDUCATION. *Inventory of Beliefs: Instructor's manual.* Washington, D. C.: ACE, 1950.

AMERICAN COUNCIL ON EDUCATION. *Test of Critical Thinking: Instructor's manual.* Washington, D. C.: ACE, 1953.

BROWN, D. R., & BYSTRYN, DENISE. College environment, personality, and social ideology of three ethnic groups. *J. soc. Psychol.,* 1956, **44**, 279-288.

CARLSON, E. R. Attitude change through modification of attitude structure. *J. abnorm. soc. Psychol.,* 1956, **52**, 256-261.

FOSTER, F. S., STANEK, R., & KRASSOWOSKI, W. *The impact of a value-oriented university on student attitudes and thinking.* Santa Clara, California: University of Santa Clara, 1961.

GOLDSEN, ROSE K., ROSENBERG, M., WILLIAMS, R. M., & SUCHMAN, E. A. *What college students think.* New York: Van Nostrand, 1960.

JACOB, P. E. *Changing values in college.* New York: Harpers, 1958.

LEHMANN, I. J., & DRESSEL, P. L. *Critical thinking, attitudes, and values in higher education.* East Lansing: Michigan State University, 1962.

LEHMANN, I. J., & IKENBERRY, S. O. *Critical thinking, attitudes, and values in higher education: A preliminary report of research.* East Lansing: Michigan State University, 1959.

MCCONNELL, T. R. Change in scores on the psychological examination of the American Council on Education from the freshman to senior year. *J. educ. Psychol.,* 1934, **25**, 66–69..

MCGUIGAN, F. J. Psychological changes related to intercultural experiences. *Psychol. Rep.,* 1958, **4**, 55–60.

MATTESON, R. W. Experience-interest changes in students. *J. counsel. Psychol.,* 1955, **2**, 113–121.

MURPHY, LOIS, & RAUSHENBUSH, ESTHER. *Achievement in the college years.* New York: Harper, 1960.

NELSON, H., BLAKE, R., MORTON, JANE, & OLMSTEAD, J. Attitudes as adjustments to stimulus, background, and residual factors. *J. abnorm. soc. Psychol.,* 1956, **52**, 314–322.

NEWCOMB, T. M. *Personality and social change.* New York: Dryden, 1943.

PLANT, W. T. *Personality changes associated with a college education.* San Jose, California: San Jose State College, 1962.

PRINCE, R. A study of the relationship between individual values and administrative effectiveness in the school situation. Unpublished doctoral dissertation, University of Chicago, 1957.

ROKEACH, M. *The open and closed mind.* New York: Basic Books, 1960.

ROSENBERG, M. Psychological depression and educational attitudes. *Student Med.,* 1956, **5**, 5–20.

SANFORD, N. Personality development during the college years. *J. soc. Issues,* 1956, **12**, 3–70.

SMITH, H. P. Do intercultural experiences affect attitudes? *J. abnorm. soc. Psychol.,* 1955, **51**, 469–477.

STRONG, E. K. *Vocational interests of men and women.* Stanford, California: Stanford Univer. Press, 1943.

WEBSTER, H. Changes in attitudes during college. *J. educ. Psychol.,* 1958, **49**, 109–117.

(Received March 11, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 6, 316–321

# RELATION BETWEEN WORD-RECOGNITION ERRORS AND HAND-EYE PREFERENCE IN PRESCHOOL CHILDREN [1]

SIEGMAR MUEHL
*College of Education, University of Iowa*

Handedness and eyedness tests were administered to 62 preschool children at 2 age levels. Evidence for visual orienting behavior was obtained from a 20-item word-recognition task. Ss matched each stimulus word with 4 response choices designed to elicit a correct response, or a left, middle, or right recognition error. Ss in both age groups made more correct responses than any type of recognition error. There were no differences among types of recognition errors. Correct responses increased with age; left errors decreased more than other errors. Ss classified as left lateral, either in hand, eye, or both, made more combined left and right errors than did right hand-eye Ss. The findings considered with related research provided additional evidence that left laterality in prereaders is associated with unique patterns of perceptual behavior.

Word recognition for the beginning reader depends upon his learning a sound-to-spelling relationship. The basis for this relationship requires not only discrimination of the letters which form a word, but also orienting to these letters in a left-right order.

Teaching children this left-right orienting behavior is part of the instruction program in beginning reading. Two questions pertinent to this instructional task are: Do prereaders, when presented with a stimulus pattern involving a printed word, have any tendency to orient either to the left, right, or middle of the word? Does this tendency, if it exists, relate to hand-eye preference?

Gates and Boeker (1923) and Meek (1925) investigated the first of these questions, using nursery school and kindergarten children. When word-recognition errors and letter similarities among words were examined, both studies found the most frequent error associated with end-letter similarities among the words. Meek reported the least number associated with middle-letter similarities. None of these findings, however, was tested for reliability.

The second question has not been directly studied. Stevenson and Robinson (1953) investigated the relation between hand-eye preference and directional tendencies in kindergarten children prior to reading instruction. Using a variety of tasks to measure reversals, they found a significantly greater tendency for right-handed, right-eyed children to arrange a series of pictures in a left-right direction as compared to a right-left direction for right-handed, left-eyed children.

The purpose of the present study was further to investigate visual orienting behavior as it relates to word stimuli and hand-eye preferences in prereaders.

## METHOD

### *Handedness Tests*

Tests for hand preference were adapted from Updegraff (1932). Preliminary testing was done to select items eliciting unimanual and hand-consistent responses. The final test consisted of six items including: coloring with a crayon, cutting paper, hammering, building a block bridge, spooning marbles into a cup, and throwing a ball. Handedness scores ranged from 0 to 6, with 0 indicating consistent left handedness, and 6 indicating consistent right handedness.

---

[1] The author wishes to thank S. Kreitzer for his role in testing the children and for his many useful suggestions in planning the experimental procedures.

Short- and long-term reliability of the handedness test was estimated from preliminary testing. For 23 preschool subjects tested and retested from 1 to 16 days (mean of 12 days), the mean change in score was .65 with a range from 0 to 3 points. Twenty-one subjects changed 1 point or less. For 13 subjects tested and retested at a 10-month interval, the mean change was 1.01 with a range from 0–3 points. Nine subjects changed 1 point or less.

## *Eyedness Test*

Three objective tests for eyedness were adapted from Hildreth (1950). Two of the tests required the subject to sight through a 7×8 inch board with a 3-centimeter hole cut in the middle. This sighting board was mounted for vertical adjustment. In the first test, the subject was seated in a chair at arm's length from the board. The sighting board was attached to the end of a table and adjusted to the subject's height. The experimenter then sat at the opposite end of the table about 4 feet from the subject, exposed a small black and white picture, and asked the subject to look through the hole in the board and tell what he saw. In the second test the sighting board was adjusted to the subject's standing height. The subject was instructed to walk up to the board, look through the hole, and report what the picture was. The third eyedness test required the subject to sight through Miles' (1930) V scope. In each of the three eyedness tests, three sighting trials were given, using different pictures. The experimenter observed which eye the subject used in sighting in each case. Eyedness scores ranged from 0 to 9, with 0 indicating consistent left eyedness and 9 indicating consistent right eyedness.

Short- and long-term reliability of the eyedness test was estimated from the preliminary testing. For 23 preschool subjects tested and retested from 1 to 16 days (mean of 12 days), the mean change in score was .74 with a range from 0 to 3 points. Nineteen subjects changed 1 point or less between testings. For 13 subjects tested and retested at a 10-month interval, the mean change was 1.23 with a range from 0 to 5 points. Nine subjects changed 1 point or less; only one subject changed as much as 5 points.

## *Word-Recognition Task*

The apparatus for presenting the word stimuli consisted of a 2×2 inch slide projector and an 8½×11 inch white translucent glass screen mounted at the end of a box tunnel 2½ feet in length. The projector was located at the end of the box tunnel to the rear of the glass screen. By this means the projector light was confined to the tunnel and the image projected on the back of the screen. When the subject faced the screen, the word stimuli appeared white on a black background. The screen itself was so masked as to present two openings. Centered at the top was a 2½-inch square opening where a single stimulus word appeared. At the bottom of the screen was a 1½×11 inch opening where four response words appeared. The four words were equally spaced across the long dimension. The stimulus and response words were presented simultaneously. Each word measured on the screen about 20 millimeters in length and 7–10 millimeters in height. Directly in front of each response word was a red button. The button served to indicate a discrete response by the subject, and, in the case of a correct choice, appeared to make the stimulus words disappear from the screen. Actually the slide projector was activated by the experimenter using a hidden control.

The words used in the recognition task were typed in lower case and mounted on cards to be photographed for slides. A total of 24 word sets was prepared. A word set consisted of a four-letter stimulus word, mounted at the top of the card, and four response words mounted at the bottom. Three of the response words contained rearrangements of the letters in the stimulus word. The letter rearrangements were designed to detect visual orienting responses directed to letter sequences either at the left, middle, or right of the stimulus and response words, to the exclusion of the remaining letters. The fourth response word contained the same letter arrangement as the stimulus word. Thus, when "girl" appeared as a stimulus word, the response words were: "gilr," "lirg," "igrl," and "girl." The choice of "gilr" indicated a left-orienting response and was scored as a left-recognition error; "lirg," a middle orienting response scored as a middle recognition error; "igrl," a right-orienting response scored as a right recognition error. The choice of "girl" was scored as a correct recognition response.

The 24 word sets were divided into 6 blocks of 4 sets each. The positions of the 4 types of response words were counterbalanced within each block so that each type appeared in all

4 positions. One of the 6 blocks is illustrated in Table 1.

In addition to the word sets, 4 sets were prepared containing familiar geometric forms including a square, a rectangle, a circle, and a triangle. Each form served as a stimulus and a response. Counterbalancing the responses was similar to the word sets.

The word-recognition task was preceded by two preliminary tasks. In the first, called the instruction-learning task, the 4 sets of geometric forms were used. In the second, called the word warm-up task, 4 word sets were used. The main recognition task consisted of the remaining 20 word sets.

When the subject was seated in front of the glass screen, the first geometric form set was presented. The subject was instructed (*a*) to look at and point to the form appearing in the top opening of the screen, (*b*) to look below and find a picture that looked the same, and (*c*) to push the button in front of that picture. If the subject's response was correct the picture disappeared. If incorrect, the experimenter pointed to the correct picture and pressed the button. In addition, for the first two sets, the experimenter pointed to each one of the responses, asking the subject each time whether this picture was the same as the one above. The subject was shown the four geometric sets twice. A criterion of six of eight correct responses was required to continue. No subjects were dropped for failure to reach this criterion in the instruction-learning task. The word warm-up task followed immediately. The instructions for the first two word sets were identical to those used with the geometric forms except "word" was substituted for "picture." On completion of the fourth warm-up set the subject was told he would see some more words and was to try to see how many he could get right. Throughout this task the subject was given verbal encouragement and reminded to look at all the words before pressing the button. The primary incentive, however, was to make the words disappear. The subject received a toy on completing the task.

With most subjects, the handedness and eyedness tests and the word-recognition task were administered in the same session; the session required was 20–30 minutes.

TABLE 1

SAMPLE BLOCK OF RESPONSE WORDS COUNTERBALANCED FOR PRESENTATION POSITION

| Button position | | | |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| gilr | igrl | lirg | girl |
| look | kool | olok | loko |
| ujmp | jupm | jump | pumj |
| eosr | rose | roes | orse |

### *Subjects*

The subjects were 62 children from the University and Parent's Cooperative Preschools at the University of Iowa. Four additional subjects were eliminated: two because they could read, and two because they were ambidextrous with respect to eye preference. The subjects were divided into two age groups based on their preschool group assignments. Group CA IV consisted of 23 subjects from the younger groups with a mean age of 48.7 months, and a range from 43 to 54 months. Group CA V consisted of 39 subjects from the older groups with a mean age of 58.2 months, and a range from 53 to 63 months.

## RESULTS

### *Handedness-Eyedness Tests*

Four different hand-eye groups were constituted from the results of the handedness and eyedness tests. These groups are shown in Table 2. Only subjects giving evidence of a preferred hand or eye were included.

On the handedness test, subjects scoring from 0 to 2 were classed as left handed and from 4 to 6 as right handed. The distribution of handedness was markedly bimodal with the majority of

TABLE 2

NUMBER OF SUBJECTS IN HAND-EYE GROUPS

| Hand-eye | Age groups | | |
|---|---|---|---|
| | CA IV | CA V | Combined |
| Right-right (RR) | 10 | 24 | 34 |
| Right-left (RL) | 6 | 8 | 14 |
| Left-left (LL) | 2 | 4 | 6 |
| Left-right (LR) | 5 | 3 | 8 |
| Total | | | 62 |

subjects scoring either 0–1 or 5–6. On the eyedness test, subjects scoring 0–3 were classed as left eyed and from 6–9 as right eyed. The distribution of these scores was also bimodal, with the majority of subjects scoring either 0–1 or 8–9.

To assess the effect of handedness and eyedness on type of word recognition errors, the groups appearing in Table 2 were variously combined to make six categories of pairs for purposes of comparison. These categories are described in Table 3.

### *Word-Recognition Task*

Each of the six pairs of hand-eye groups shown in Table 3 was compared by means of a Lindquist (1953) Type VI analysis of variance. To obtain age and hand-eye group proportionality, it was necessary in some instances to drop subjects on a random basis (see Tables 2 and 3). In the analyses, age and hand-eye groups within each category were compared for all pairs of response choices, making a total of six comparisons.[2] For example, in the right-right versus right-left category, the six analyses included comparisons between: (*a*) correct responses and left recognition errors, (*b*) correct and middle, (*c*) correct and right, (*d*) left and middle, (*e*) left and right, and (*f*) middle and right recognition errors. The main effects in each of these analyses were: response choice as the within factor; age and hand-eye groups as the between factors. The findings are summarized under the main effect headings. Response choice and age data were based on Category 3 (see Table 3) analyses for the maximum number of subjects.

TABLE 3

HAND-EYE GROUPS COMPARED IN THE VARIOUS ANALYSES

| Comparison categories | Laterality characteristics |
|---|---|
| 1. RR versus RL<br>(21)[a] (14)[a] | Consistent versus mixed hand-eye |
| 2. RR+LL versus RL+LR<br>(24) (22) | |
| 3. RR+LR versus LL+RL<br>(35) (20) | Right eye versus left eye |
| 4. RR+RL versus LL+LR<br>(32) (14) | Right hand versus left hand |
| 5. RR versus LL<br>(30) (6) | Consistent right versus consistent left hand-eye |
| 6. RR versus RL+LR+LL<br>(25) (25) | Consistent right versus left, all combinations |

[a] Number of *Ss* in groups when combined proportionately according to age.

### *Response Choice*

Table 4, fourth column, shows the means and standard deviations for the four response choices in the word-recognition task for the combined age groups. Comparisons among the response choices showed a significant difference between correct responses and each of the three kinds of recognition errors ($p < .001$). Differences among left, middle, and right recognition errors were not significant ($p > .05$).

### *Age*

There was a significant interaction between age and type of response when correct responses were compared with left ($p < .001$), middle, and right word-recognition errors ($p < .05$). The trend of this interaction is shown in Table 4, Columns 2 and 3. Thus, comparing correct responses and left errors for the two age levels, it is apparent that correct responses increased with age and left errors decreased. The interaction pat-

---

[2] The analysis by pairs of response choices was considered appropriate since these measures were not independent.

TABLE 4

RESPONSE CHOICES IN THE WORD RECOGNITION TASK BY AGE GROUPS[a]

| Response choice | CA IV ($N = 22$) | CA V ($N = 33$) | Combined ($N = 55$) |
|---|---|---|---|
| Correct | | | |
| *M* | 7.05 | 10.15 | 8.91 |
| *SD* | 3.52 | 4.02 | 4.12 |
| Left error | | | |
| *M* | 5.09 | 3.48 | 4.13 |
| *SD* | 3.37 | 1.09 | 2.63 |
| Middle error | | | |
| *M* | 3.77 | 2.79 | 3.18 |
| *SD* | 2.04 | 2.31 | 2.26 |
| Right error | | | |
| *M* | 4.05 | 3.58 | 3.76 |
| *S.D.* | 2.03 | 2.34 | 2.23 |

[a] Data from Category 3 analysis for maximum *N*.

tern was similar but not as marked when correct responses were compared with middle and right errors.

### *Hand-eye groups*

Comparisons between hand-eye groups in the six categories described in Table 3 showed two significant differences. When right-right subjects were compared with left-left subjects, age collapsed (Category 5), it was found that the left-left subjects made significantly more combined left and right recognition errors ($p<.05$). The mean combined errors for the right-right and left-left groups were 6.90 and 10.42; the standard deviations were 3.23 and 2.77, respectively.

Similar results occurred when right-right subjects were compared with subjects having left-laterality of either hand, eye, or both (Category 6). The mean combined left and right errors were 6.80 and 8.92; the standard deviations were 3.28 and 2.98, respectively. The difference was significant ($p<.025$). In category 6, the sex distribution showed 8 boys and 17 girls to be right-right and 14 boys and 11 girls to be either right-left, left-right, or left-left. A chi square test, using Yates correction, was not significant.

## DISCUSSION

The purpose of the present study was to investigate visual orienting behavior and its relation to hand-eye preference in prereaders. Orienting behavior was indicated by type of recognition errors made in a word-matching task. Age differences among the preschool children made age an additional factor.

The findings with respect to word-recognition errors did not indicate an overall tendency for prereaders to make more left than right errors. The observed tendency to make fewer recognition errors based on middle letter similarities, though not significant, was similar to Meek's (1925) findings. The significant increase in correct responses for the 5-year group suggests that the older child has learned, to a greater degree, the habit of orienting to a total stimulus pattern before responding to it. Whether this more efficient orienting behavior is accompanied by a left-to-right pattern of viewing word stimuli was not clearly measured by the present results. The marked decrease in left errors accompanying the increase of correct responses for the 5-year group may indicate that when the 5-year-old orients to the beginning of a word he follows this initial orientation with a left-to-right viewing response. This directional aspect of the response could be accounted for by the older child's greater familiarity with picture stories in books and "funny papers," both of which require a left-to-right viewing to follow the story sequence.

The finding that left-lateral subjects made more left *and* right recognition errors than consistent right-lateral subjects, although indicating no link between laterality and a particular directional viewing tendency, does indicate a lag in learning to orient and/or respond to the total word pattern for the left-lateral subjects. When considered with the results reported by Stevenson and Robinson (1953), this finding provides consistent evidence that left laterality in prereaders is associated with unique patterns of visual or perceptual behavior. These patterns could affect word recognition efficiency in later reading. Although Stevenson and Robinson (1953) reported that their right-hand, left-eye group had no problems in learning to read, these children were considerably above average in intelligence. Children combining left laterality with average or below average intelligence may have more difficulty in learning the correct visual and perceptual habits.

In this connection, it is interesting to note the observed tendency in the present study for a greater proportion of boys than girls to be left lateral, a finding consistent with data reported by Downey (1922) based on a comparable age group. Considering the fact that girls do better than boys in reading achievement (Smith & Dechant, 1961, p. 91) this finding suggests a sex-linked relation between early left laterality, retarded orienting or perceptual behavior, and subsequent reading achievement.

A follow-up study of the laterality status and reading achievement of children used in the present study is planned.

## REFERENCES

DOWNEY, JUNE E. Dextrality types and the preschool child. *Yearb. Nat. Soc. Stud. Educ.*, 1922, **27**, 153–158.

GATES, A. I., & BOEKER, ELOISE. A study of initial stages in reading by preschool children. *Teachers Coll. Rec.*, 1923, **24**, 469–488.

HILDRETH, GERTRUDE. The development and training of hand dominance: IV. Developmental problems associated with handedness. *J. genet. Psychol.*, 1950, **76**, 39–100.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

MEEK, LOIS H. A study of learning and retention in young children. *Teach. Coll. Contr. Educ.*, 1925, No. 164.

MILES, W. R. Ocular dominance in human adults. *J. gen. Psychol.*, 1930, **3**, 412–430.

SMITH, H. P., & DECHANT, E. V. *Psychology in teaching reading.* Englewood Cliffs, N. J.: Prentice-Hall, 1961.

STEVENSON, LILLIAN, & ROBINSON, HELEN. Eye-hand preference, reversals and reading progress. In Helen Robinson (Ed.), Clinical studies in Reading II. *Suppl. educ. Monogr.*, 1953, **77**, 83–88.

UPDEGRAFF, RUTH. Preferential handedness in young children. *J. exp. Educ.*, 1932, **1**, 134–139.

(Received March 14, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 6, 322–330

# COMPARISON OF ADOLESCENT AND ADULT BEHAVIORAL NORM PROPERTIES[1]

ROBERT D. STROM
*Department of Education, Purdue University*[2]

Responses of 303 high school seniors and their 30 teachers to a questionnaire concerning 6 behavioral areas served as data to determine whether adolescents and adults differ significantly in behavioral norms and whether certain types of adolescents are more ready to accept adult norms than others. Subsequent to selection of adolescent subpopulations (based on leadership and ascendance-submission), a return potential curve was employed to plot distribution of group feeling and serve as computational base for 5 behavioral properties. The chi square, t test, and F test revealed some significant differences between teacher and student norms but the discrepancies could not be attributed to the subpopulation variables.

In the past few years sociologists, psychologists, philosophers, and educators have expressed concern over the problem termed by Coleman (1961) as, "the emergence of an adolescent subculture in an industrial society [p. 3]." Coleman refers to the changes in American culture which have served to bring about the separation of teenagers from the adult society, thereby encouraging youngsters to seek more and more peer approval for their behavior. Consequently, teenagers within their subculture have developed norms of behavior which are believed to be quite divergent from those held by adult society.

The purpose of this study was to investigate the norms of behavior for the adolescent subculture. The author sought to determine: (*a*) whether adolescents differ significantly on behavioral norm properties from adults, and (*b*) whether certain types of adolescents are more ready to accept adult norms than others.

## METHOD

### *Subjects*

As a consequence of the complex adult society, no individual or group can represent all of its facets, but it is generally accepted by adults that the teacher does at least a competent job of communicating the important values and standards of adults to adolescents. Therefore, assuming that the behavioral norms of teachers are somewhat representative of the standards of adult society, it appears logical to compare the norms of behavior for adolescents with those of their teachers in order to determine the extent this group of adults and adolescents agree. The adolescent sample of 303 seniors from Edsel Ford High School in Dearborn, Michigan, was to be compared in its norms of behavior with 30 teachers who represented adult society.

### *Subpopulations*

Since some writers (Bestor, 1955; Rickover, 1960) insist that certain types of adolescents are more like adults than others in their norms of behavior, it was deemed desirable to consider these types within the sample for separate comparison to adults. Consideration was given to the assertion that students who are recognized by classmates as being leaders and those whose personalities are ascendant rather than permissive would correspond more closely to adults in their behavioral norms than nonleaders or submissive adolescents. Leadership and ascendance-submission became the variables in selecting the adolescent subpopulations. It was assumed that the ascendant person may be a leader or a nonleader; similarly, a submissive person may be a leader or a nonleader. Therefore, each individual was to be assigned to one of four subpopulations, namely: (*a*) leader ascendant, (*b*) nonleader ascendant, (*c*) leader submissive, or (*d*) nonleader submissive.

---

[1] The present research was conducted under the direction of A. Zander, Director of social research at the University of Michigan, and H. Bretsch, Assistant Dean of the graduate school at the University of Michigan.

The author gratefully acknowledges the help and encouragement of G. W. Allport of Harvard University.

[2] Now at the National Education Association, Washington, D. C.

### *Desirability of Properties*

The mere knowledge about agreement or disagreement is superficial if we cannot know in what ways and under what conditions adults and adolescents agree or disagree. In the absence of some scheme for representing the structure of a norm, much of its potential as a tool for analysis of groups fails to be realized. However, using a statistical tool called a "return potential curve" (Jackson, 1960) the essential properties incorporated in the idea of "norm" may be derived. The return curve enables the investigator not only to determine the extent to which teachers and adolescents differ in their norms of behavior, but also: (*a*) the limits of behavior which each group report they will approve or tolerate; (*b*) the intensity of approval or disapproval they state is evoked by certain types of behavior; (*c*) the proportion of the possible range of behavior that would be in their own self-report, approved; (*d*) the point of ideal behavior as perceived by each group; and (*e*) the relative degree of crystallization for the norms of behavior of teachers and students. Hence the investigation concerned comparison of norm properties rather than merely norms themselves.

### *Behavioral Areas*

Subjects in the study were questioned on six areas of behavior. Norms examined were of two basic types, namely, those which concern student behaviors that teachers might probably have direct influence on and those behaviors over which teacher influence is probably less direct. The more direct behaviors were (*a*) contradiction of teachers, (*b*) homework time, and (*c*) class discussion participation. The less direct areas of influence considered were (*a*) dating, (*b*) clique deviance, and (*c*) physical education participation.

A scale of approval-disapproval feelings, ranging from numeral notation of 4 for strongly approving to −4 for strongly disapproving through a midpoint of indifference at 0, was provided the respondents for their judgment on the eight frequencies of the behavioral dimension for each question. A sample question from the investigation follows.

> Although there are no societal rules governing the frequency of dating, students have ideas concerning how often they should date while in school. For example, students who date only once a year may find their dating behavior disapproved by some classmates. On the other hand, students who date every evening may also be considered for disapproval by some classmates. If a student dates in the frequencies mentioned below, give your strength of approval or disapproval with the scale number of your choice in the column to the right.
>
> (a) Three times a year —
> (b) Once a month —
> (c) Twice a month —
> (d) Once a week —
> (e) Twice a week —
> (f) Three times a week —
> (g) Five times a week —
> (h) Daily —

### *Major Instrument*

The properties of a norm are derived from a return potential curve as shown in each of Figures 1–6. The horizontal axis is the dimension of behavior along which the frequency of action is cited while a dimension of evaluation is represented by the vertical axis. For any particular behavior dimension, the amount of approval or disapproval felt by members of a group toward a particular act may in principle fall anywhere along the evaluation dimension from strongly approving to strongly disapproving through some middle point of indifference. It is possible then in any concrete situation to plot a curve describing the feelings of the group members. Approved behavior would be revealed by positive potential return above the point of indifference whereas negative return characterizes behavioral frequencies which are disapproved.

### *Determination of Subpopulation*

The seniors in their 12 separate homerooms were administered a leadership index. The names of all members of the particular homeroom were listed and after each name appeared

TABLE 1

NUMBER OF SUBJECTS IN EACH SUBPOPULATION

| Group | Men | Women |
|---|---|---|
| Ascendant leaders | 48 | 51 |
| Ascendant nonleaders | 28 | 22 |
| Submissive leaders | 39 | 14 |
| Submissive nonleaders | 56 | 45 |
| Total | 171 | 132 |

a series of numerals from 1 to 5. Students were directed to circle the numeral after each student's name which would indicate to what extent the rater considered the ratee to be a leader. The numeral 5 indicated that the rater believed the ratee to be outstanding in leadership whereas a rating of 1 indicated that the ratee was not a leader. After the subjects had been rated by their peers, the subjects' leadership indexes were computed by multiplying the circled leadership numerals by the frequencies in which each number occurred and dividing this product by the number of students who did the rating.

The Allports' (1928) *Ascendance Reaction Study* was used to assess the relative ascendance or submission in the personality of the adolescent respondents. A fairly large number of situations is verbally presented in the reaction study in which the respondent selects one of a group of standardized responses which most nearly characterizes his usual behavior in that situation. Each response choice is considered as having a diagnostic value for indicating ascendance (plus scores) or submission (minus scores). The final ascendance-submission score is the algebraic sum of scores for separate items.

From a frequency distribution the median score in the leadership index was chosen as the separation for leaders and nonleaders and the ascendance-submission scores exceeding 0 in a plus direction categorized the ascendant persons. Individuals were then placed into one of the four subgroups depending on their scores, as shown in Table 1.

## RESULTS

Before comparing the degree that various properties were assigned to the norms, it is useful to examine Figures 1–6, which contain the return potential curves for all subpopulations on all questions. Note that the teachers render much stronger approval than students at the lower end of the frequency range and stronger disapproval at the greater frequency end of the behavior dimension in all norms except that in Figure 5 (homework). In Figure 1 (dating) and Figure 5 (homework) teachers and students seem to agree somewhat but in Figure 4 (athletics), complete disagreement is evident. Throughout Figures 1–6, it is seen that teachers evidence much stronger feelings of approval-disapproval than do their students.

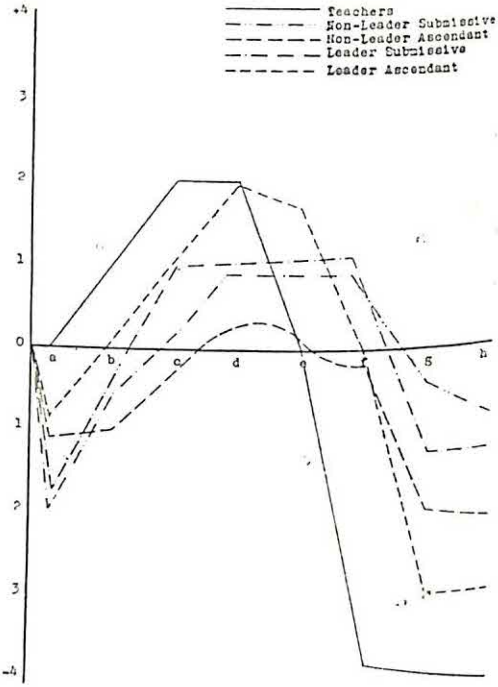


FIG. 1. Dating norm curve for all subpopulations and teachers. (a. Three times a year; b. Once a month; c. Twice a month; d. Once a week; e. Twice a week; f. Three times a week; g. Five times a week; h. Daily.)

Subsequent to surveying differences in return potential curve structures between teachers and students, the comparative properties were computed for teacher and student subpopulations by summation of feeling on all questions,

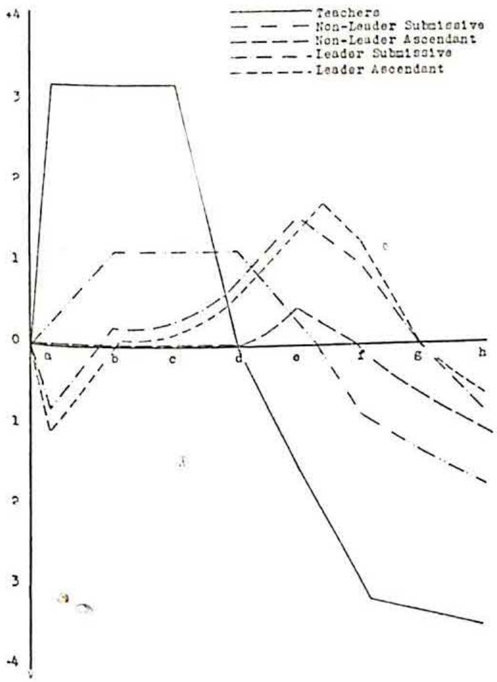


FIG. 2. Contradiction norm curve for all subpopulations and teachers. (a. Once a semester; b. Once a month; c. Twice a month; d. Three times a month; e. Once a week; f. Twice a week; g. Three times a week; h. Four times a week.)

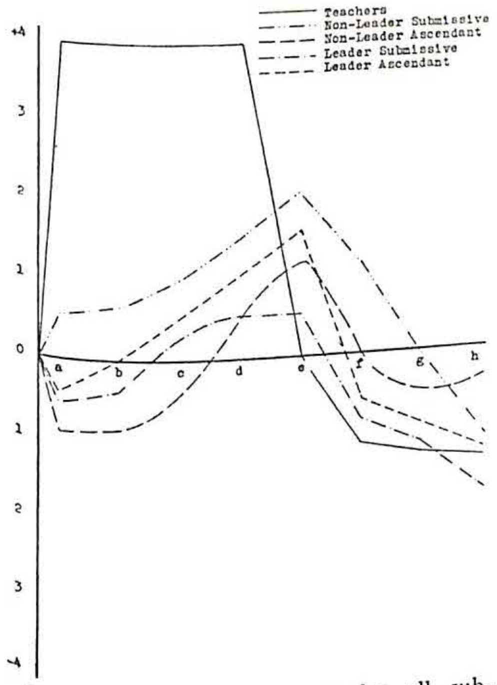


FIG. 3. Clique norm curve for all subpopulations and teachers. (a. Once a year; b. Twice a year; c. Three times a year; d. Four times a year; e. Once a month; f. Twice a month; g. Once a week; h. Twice a week.)

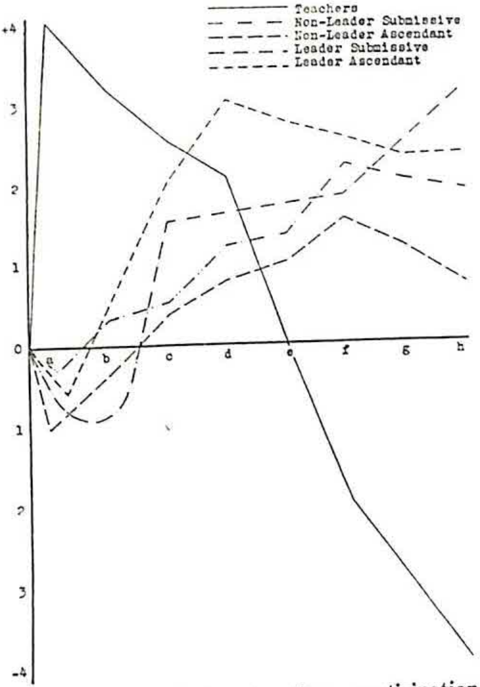


FIG. 4. Physical education participation norm curve for all subpopulations and teachers. (a. Once a semester; b. Once a month; c. Twice a month; d. Three times a month; e. Once a week; f. Twice a week; g. Three times a week; h. Four times a week.)

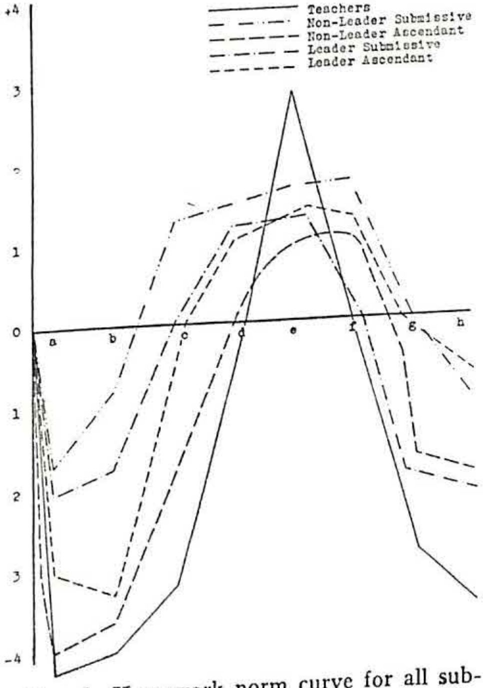


FIG. 5. Homework norm curve for all subpopulations and teachers. (a. 1 hour a week; b. 2 hours a week; c. 3 hours a week; d. 6 hours a week; e. 12 hours a week; f. 18 hours a week; g. 24 hours a week; h. 30 hours a week.)

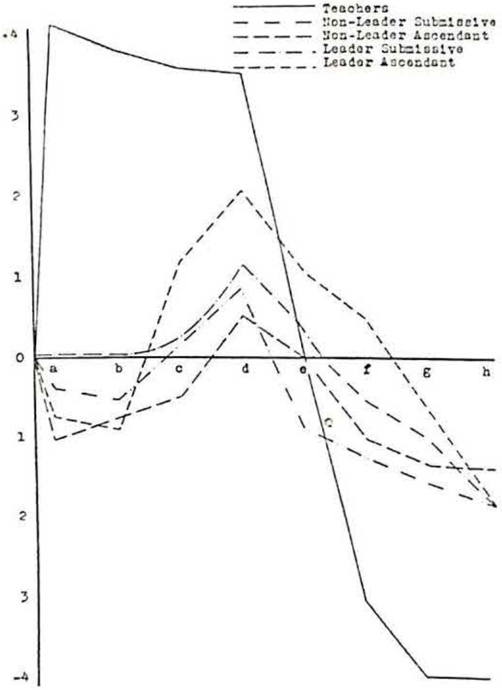


FIG. 6. Class discussion norm curves for all subpopulations and teachers. (a. Once a week; b. Twice a week; c. Once an hour; d. Twice an hour; e. 3 times an hour; f. 5 times an hour; g. 7 times an hour; h. 10 times an hour.)

yielding a composite score which served as an index for the tests of significance. Index scores were used instead of one norm at a time as we assumed the aggregate score, representing a more general attitude about all norms among both students and teachers, was more relevant to our major interest than a detailed study of each separate norm.

"Range of tolerability," which is the part of a behavior dimension members of a group approve, was measured by taking the mean number of positive positions over the behavior dimension for all six norms as the tolerable score for each individual. Tolerable scores were then divided into categories for chi square comparison. Table 2 shows that the range of tolerability for students does not differ significantly at the .05 level from that of the teachers. Also, there is no significant difference between the boys and girls in their ranges of tolerability.

The "intensity" of individual norms was measured by a summation of curve ordinates at each position on the behavioral dimension. Mean intensity of each question was computed for each group and then the mean of the six norms, as shown in Table 3, was used for comparison using the *t* test procedure. It is evident in Table 3 that there is a significant difference at the .01 level between the strength of feelings held by teachers and students on the intensity of the norms. The mean intensity of boys was slightly higher than that of girls but not significantly so.

Table 4 shows a comparison of "approval-disapproval ratios" between teachers and students. This property, which indicates the extent behavior is sanctioned by a group, is computed by

TABLE 2

RANGE OF TOLERABLE BEHAVIOR FOR ALL NORMS—TEACHERS VERSUS STUDENTS

| Sample | Tolerable Score 1–3 | Tolerable Score 4–5 | Tolerable Score 6 + | $\chi^2$ [a] Teachers versus students |
|---|---|---|---|---|
| Teachers | 5 | 22 | 3 | — |
| All students | 135 | 135 | 33 | 6.6663 |
| Boys | 73 | 75 | 23 | 6.6636 |
| Girls | 62 | 60 | 10 | 6.6691 |

[a] All comparisons are nonsignificant.

TABLE 3

MEAN INTENSITY FOR ALL NORMS—TEACHERS VERSUS STUDENTS

| Sample | *M* intensity | *t* of difference for teachers versus students |
|---|---|---|
| Teachers | 18.1 | — |
| All students | 8.0 | 2.83* |
| Boys | 8.0 | 3.02* |
| Girls | 7.9 | 2.63* |

* $p < .01$.

TABLE 4

APPROVAL-DISAPPROVAL RATIO FOR ALL NORMS—TEACHERS VERSUS STUDENTS

| Sample | *M* ratio | *t* of difference for teachers versus students [a] |
|---|---|---|
| Teachers | .9 | — |
| All students | 1.3 | 1.61 |
| Boys | 1.2 | 1.46 |
| Girls | 1.5 | 1.76 |

[a] All comparisons are nonsignificant.

dividing the mean intensity for positive return by the mean negative potential over the entire behavior dimension. Mean ratios for each question were computed for each group and then the mean of the six norms served as an index for comparison by *t* test. No difference at the .05 level was found. The boys seemed to sanction the behaviors less on the whole than did the girls but the difference was not statistically significant.

That behavior which represents the ideal in the eyes of group members is called the "point of maximum return." Computation of the mean maximal point for each student was determined after which individuals were placed in one of three groups for chi square comparison. Table 5 indicates that both the total class and the boys were significantly different in point of maximum return from the teachers. However, the girls were not significantly different from the teachers. Thus, in the property of maximum return, girls appear to conform more to the norms of teachers than do boys.

The crystallization or, conversely, ambiguity of a norm, drawn from the variance of the return potential curve, indicates to what extent members of a group have really established a norm or made up their minds. The mean variance of each group was computed for each question after which the mean of the means was determined for comparison by the *F* test method. In the cases of boys, girls, and the total class, Table 6 shows significance at the .01 level for the degree of crystallization between student norms and those of teachers.

At this juncture teachers were com-

TABLE 5

MAXIMUM RETURN FOR ALL NORMS—TEACHERS VERSUS STUDENTS

| Sample | Maximum Points A–B | Maximum Points C–D | Maximum Points E+ | $\chi^2$ Teachers versus students |
|---|---|---|---|---|
| Teachers | 15 | 10 | 5 | — |
| All students | 53 | 151 | 99 | 8.3600* |
| Boys | 29 | 86 | 56 | 9.2825* |
| Girls | 24 | 65 | 43 | 7.4375 |

* $p < .05$.

TABLE 6

DEGREE OF CRYSTALLIZATION FOR ALL NORMS—TEACHERS VERSUS STUDENTS

| Sample | *M* variance | *F* of difference for teachers versus students |
|---|---|---|
| Teachers | 5.0 | — |
| All students | 13.3 | 2.66* |
| Boys | 14.1 | 2.82* |
| Girls | 12.5 | 2.51* |

* $p < .01$.

pared to the subpopulations, and to the subgroups (separate sexes of subpopulations). Moreover, the subgroups were compared with each other. The measures of comparison were the same as have been illustrated in Tables 2–6. Statistical differences for the properties of tolerable range and maximal point were found using the chi square method; intensity and approval-disapproval ratio differences determined by the *t* test; and the degree of crystallization property compared by the *F* test.

The major findings of this study show that high school seniors do differ significantly on certain properties in the norms of behavior from their teachers. The students as a group deviate from teacher behavioral norms on the properties of intensity, maximum return, and degree of crystallization. Statistical dissonance from teachers on the intensity property indicates that students do not feel as strongly about the behaviors studied as teachers do and therefore would not be as concerned about how their peers acted on these behaviors as would the teachers. The discrepancy in maximum return shows students and teachers differ in their perceptions of just what behavior is to be considered as ideal, indicating a behavioral goal conflict. Moreover, students do not have norms of behavior which are as crystallized or developed as norms of teachers. Thus, on those behaviors studied the adolescents had not made up their minds to the extent that teachers had concerning what was to be approved or disapproved.

On the other hand, the adolescents in their norm properties of tolerable range and approval-disapproval ratio were not statistically different from their teachers. These results seem to indicate that the restrictions as to what behavior will be allowed or tolerated is nearly the same for adolescents as teachers. In addition, the proportion of possible range of behavior that would be approved or disapproved by the adolescents is in accord with the ratio accepted by the teachers.

The answer as to why certain discrepancies exist between the properties of teacher and student norms is not to be found by asserting that too few adolescents are ascendant in personality. Approximately half of the student sample was ascendant. Comparing these ascendant adolescents versus their submissive peers (ignoring sex and leadership) with teachers, it was found that ascendance does not seem to be an important determinant on most norm properties. Just as we cannot attribute the cause of statistical differences between adult-adolescent norms to the variable of ascendance, we cannot attach the cause to leadership. Leaders versus nonleaders (ignoring sex and ascendance scores) were compared with teachers and the findings indicate that leadership was not an important factor in determining the extent to which adolescents will differ from teachers. Similarly, girls versus boys (ignoring leadership and ascendance) were compared with the result that sex was declared unimportant as a determinant.

The possibility that certain of the subgroups would be more like teachers than other subgroups led the investigator to determine which of the eight subgroups (including boys and girls) would be most like and unlike teachers. The findings show that no one particular subgroup was in consistent agreement with teachers nor was any particular subgroup more unlike teachers than the others.

Finding the extent to which subgroups differed from one another was important in determining whether one or more of the subgroups deviated from their peers. With few exceptions, the uniformity of subgroup norms would lead one to accept as reasonable the assertion that students are mostly the same in their behavioral norm properties.

Of the three types of behavior in which teacher influence was more likely to occur, there appears to be no closer proximity to adult norms than those behaviors in which teacher influence is less likely. It would seem that direct teacher influence is of little importance in determining the behavioral norms for this group of senior students.

Some further interesting information emerging as concomitant outcomes of this study concerns the relative distribution of ascendance and submission within the adolescent subculture as well as the relationship between ascendance and leadership. This study shows that if a person is a leader the likelihood is twice as great that he will be ascendant than that he will not be. Conversely, if he is a nonleader, it seems twice as likely that he will be submissive as ascendant. The correlation between ascendance-submission scores and leadership was .28 ($p < .005$) for senior boys and .37 ($p < .005$) for senior girls.

## DISCUSSION

This study demonstrates the success of a method by which the social climate of youth may be known. Adult concern with adolescent conformity may be shown to be valid or inappropriate depending on the direction of compliance and the types of behavior it may encompass. Nevertheless, to understand the ways in which peer pressures operate, to ascertain the areas of behavior in which these pressures are most influential, and to seek ways by which such forces might be redirected for personal growth will be helpful in affecting an improvement of the educative process.

In terms of educational setting, we have always known what teachers expect of students; however, our knowledge is both recent and sparce concerning the nature and influence of student expectations upon fellow pupils. Yet, educators readily admit that both the amount and kind of learning are influenced by student climate. Indeed, to avoid rejection teenagers find it necessary to be cognizant of and in accord with peer norms regarding such academic behavior as class participation and allocation of homework time. No less expedient should it be for teachers to become acquainted with student norms lest they fail to offer an acceptable province for learning.

Evaluating the teaching of attitudes has been a task of anxious dimension for adults since time immemorial. It is now possible to measure whether desired outcomes of learning in human relations and conduct are being realized. This would encompass a host of concerns from patriotism to racial tolerance. Using dating as an area of personal relations, this effort illustrates how one can determine the kinds and limits of behavior youngsters will tolerate, the behavior perceived by a group as ideal, the intensity of approval-disapproval evoked by certain practices, and the degree to which ideas have been accepted.

Helping youngsters maintain satisfactory relations with their fellows is a difficult but necessary task of the educator. For the rejected child, something more than identification as an isolate is needed to produce adjustment. Comparing the deviant's norms of behavior with his peer group will reveal the conditions under which exclusion occurs and enable the counselor to provide direction for the student. To best foster adjustment and to increase our knowledge of personality development we need to examine behavioral norms for children of all ages. This would involve consideration of subgroups such as the creative and gifted. In the final analysis, a teacher's effectiveness is adversely affected if his understanding of student norms is not in juxtaposition to reality. It follows that to better know the student is to be a better teacher.

## REFERENCES

ALLPORT, F. H., & ALLPORT, G. W. *Ascendance reaction study*. Boston: Houghton Mifflin, 1928.

BESTOR, A. *Restoration of learning.* New York: Knopf, 1955.

COLEMAN, J. *Social climates in high school.* (Publ. No. OE33016) Washington, D. C.: United States Department of Health, Education, and Welfare, 1961.

JACKSON, J. M. Structural characteristics of norms. In National Society for the Study of Education, *The dynamics of instructional groups*. Washington, D. C.: NSSE, 1960.

RICKOVER, H. *Education for freedom.* (2nd ed.) New York: Harper, 1960.

(Received May 2, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 6, 331–336

# ROLE OF DISCRIMINABILITY IN MEANINGFUL PARALLEL LEARNING

DAVID P. AUSUBEL AND MOHAMED YOUSSEF
*Bureau of Educational Research, University of Illinois*

Previously learned relevant background knowledge (Christianity) significantly facilitated the learning and retention of the 1st of 2 parallelly related, unfamiliar learning passages (Buddhism and Zen Buddhism) when the influence of verbal ability was statistically controlled. Similarly, with verbal ability statistically controlled, knowledge of the Buddhism passage significantly facilitated the learning and retention of the Zen Buddhism passage. These facilitating effects were attributed to the increased discriminability of the new set of parallel ideas in the learning passage when previously learned related knowledge is clear and stable. The organizers had a significant facilitating effect on the learning and retention of the 1st passage, but failed to facilitate significantly the learning and retention of the 2nd passage.

---

Subject matter in the classroom is frequently organized along parallel lines. That is, related materials are presented in serial fashion, but there is no *intrinsic* sequential dependence from one topic to the next. Unlike the case in sequentially organized subject matter, successive learning tasks are inherently independent of each other in the sense that understanding of Part II material does not presuppose understanding of Part I material. Each set of material is logically self-contained and can be adequately learned by itself without any reference to the other; order of presentation is therefore immaterial. This situation, for example, prevails in presenting alternative theoretical positions in ethics, religion, and epistemology; opposing theories of biological evolution; and different systems of learning and personality theory.

Nevertheless, although successive learning tasks of parallelly organized material are not intrinsically dependent on each other, much cognitive interaction obviously occurs between them. Earlier learned elements of a parallel sequence serve an orienting and subsuming role in relation to later presented elements. The latter is comprehended and interpreted in terms of existing understandings and paradigms provided by analogous, familiar, previously learned, and already established concepts in cognitive structure. Hence, for learning of the unfamiliar new ideas to take place, they must be adequately discriminable from the established, familiar ideas; otherwise the new meanings are so permeated with ambiguities, misconceptions, and confusions as to be partially or completely nonexistent in their own right. If, for example, the learner cannot discriminate between new idea A′ and old idea A, A′ does not really exist for him; it is phenomenologically the same idea as A. Furthermore, even if the learner can discriminate between A and A′ at the moment of learning, unless the discrimination is sharp and free from ambiguity and confusion, there will be a tendency over time for A′ to be remembered as A as the two ideas interact during the retention interval. Since in any case it is characteristic of the organization of knowledge in memory for unfamiliar and unestablished ideas to be reduced to familiar and established ideas, this process of reduction is obviously accelerated if the new and unfamiliar concepts are indiscriminable from the old. If new ideas are not sufficiently discriminable from previously learned related ideas, they can be adequately represented by the latter for memorial purposes.

In this experiment the two parallel and serially presented sets of unfamiliar ideas constituting the learning tasks are the principal tenets of Buddhism and Zen Buddhism, respectively. The problem of discriminability for the learner arises first in differentiating the new Buddhism ideas from analogous established ideas about Christianity in his existing cognitive structure, and then in differentiating the Zen Buddhism ideas from the newly learned Buddhism ideas. The experimental subjects in each instance also receive the benefit of an advance organizer explicitly designed to increase discriminability by comparing the two sets of ideas in question at a high level of abstraction, generality, and inclusiveness.

It is hypothesized, first, that the discriminability of new ideas from related previously learned ideas, and hence the efficiency with which they are learned and retained, is a function of the clarity and stability of the previously learned ideas. According to this hypothesis, the discriminability of the new Buddhism ideas depends upon the stability and clarity of the learner's existing ideas about Christianity. If the Christianity ideas are vague, ambiguous, or confused, it is extremely difficult for the learner to differentiate the Buddhism ideas from them. Similarly, this hypothesis predicts that the discriminability of the Zen Buddhism ideas depends upon the stability and clarity of the analogous, newly learned Buddhism ideas.

Second, it is hypothesized that an advance organizer increases the discriminability of new material from previously learned related material, and thereby enhances its learning and retention, particularly in those subjects who either have less verbal ability or who find the new material less discriminable because of the relative instability and unclarity of the analogous, previously learned ideas in their cognitive structure.

## METHOD

### *Subjects*

The experimental population consisted of 162 predominantly senior undergraduates (95 women and 67 men) in six sections of an educational psychology course at the University of Illinois. All subjects were enrolled in 1 of 10 teacher education curricula at the secondary school level. The experiment was conducted separately in each section as a required laboratory exercise and was performed during regularly scheduled class hours. In order to maximize ego involvement, the subjects were informed that after the data were processed, their individual scores, as well as the class results, would be reported to them.

### *Learning Passages, Measuring Instruments, and Organizers*

The learning material used in this study consisted of two specially prepared 2,500-word passages dealing with the principal concepts of Buddhism and Zen Buddhism, respectively. The topic of Buddhism was chosen both because it was explicitly unfamiliar to undergraduate students and because it was parallelly related to previously learned ideas (i.e., Christian doctrines) generally familiar to all of our subjects and, presumably, reasonably well-established in the cognitive structure of most of them. Both criteria were important since the study was concerned with the effects of discriminability between unfamiliar new material and previously learned related ideas. The Zen Buddhism material was similarly chosen because of its obvious parallel relationship to the Buddhism passage. The use of unfamiliar learning material also made it possible for all subjects to start from approximately the same baseline in learning the passages.

Empirical confirmation of the unfamiliarity of the Buddhism and Zen Buddhism materials was obtained when a group of comparable subjects who had not studied the learning passages, made scores which, on the average, were only slightly and not significantly better than chance.

Knowledge of the Buddhism material was tested by a 45-item multiple-choice test with a corrected split-half reliability of .80. Knowledge of the Zen Buddhism material was measured by a 35-item multiple-choice test with a corrected split-half reliability of .81. The correlation between the two tests after study of the materials was .60. A 36-item multiple-choice test with a corrected split-half reliability of .84 was used to test the subjects' knowledge of Christianity. Items on this latter test were noncontroversial, dealing with the Old and New Testaments, Church history, and denominational beliefs and differences. Scores on all tests showed a satisfactory range of variability, and their distribution did not differ significantly from the normal distribution curve.

The experimental group studied two 500-word introductory passages (comparative organizers), the first pointing out the principal similarities and differences between Buddhist and Christian doctrines, and the second performing the same function for Buddhist and Zen Buddhist doctrines. As a control treatment procedure, the control group studied two nonideational introductory passages of a historical and biographical nature about Buddhism and Zen Buddhism. No information was included in any of the introductory passages that could be of direct help in answering questions on the Buddhism or Zen Buddhism tests. In a previous study (Ausubel & Fitzgerald, 1961), by giving the Buddhism test to a special control group which had studied the comparative organizer without the Buddhism learning passage, it had been empirically ascertained that mere exposure to the organizer per se did not constitute an independent advantage to the experimental group in taking the Buddhism test.

Scores on the verbal portion of the School and College Ability Test (SCAT), reflective of verbal ability, were available for all subjects.

### *Procedure*

On the first day of experimentation, all subjects took the Christianity test, and then (after assignment to a treatment group) studied either the comparative Christianity-Buddhism organizer or the control (historical) introductory passage for 8 minutes. Membership in a given treatment group was determined by random assignment. The population of each of the treatment groups was also stratified by sex so that the proportion of men subjects to women subjects would be the same in each group. This was necessary because of the women's significantly higher Verbal SCAT scores. It was possible to administer the two treatments simultaneously because they consisted of identical appearing introductory passages (with identical sets of directions), differing only in content. The principle of random assignment to treatment groups was relied upon to equate experimental and control groups with respect to relative learning ability and knowledge of Christianity. To equate the treatment groups with respect to prior exposure to the Buddhism and Zen Buddhism materials, those few subjects who had taken a course in comparative religion were divided equally in random fashion between the two treatment groups.

In order to control for the effects of classroom climate, instructor, and situational variables in the six sections, students *within* each section were equally divided between the two treatment groups. Since homogeneity of variance prevailed between treatment groups for the Christianity, Buddhism, and Zen Buddhism scores, it was considered justifiable to pool the scores from the various sections and to treat the scores from the two treatment groups on each of these instruments as comparable random samples from the same large population.

Two days after the first experimental session, on Day 3, both treatment groups restudied their respective introductory passages for 5 minutes and then studied the Buddhism passage for 32 minutes. The test on the Buddhism passage came 48 hours later, on Day 5. Immediately after taking the Buddhism test on Day 5, the experimental and control groups studied the comparative Buddhism-Zen Buddhism organizer and the control (historical) introduction, respectively. They had another opportunity 2 days later, on Day 7, to study their respective introductory passages for 5 minutes, after which they were given 30 minutes to learn the Zen Buddhism passage. Both groups were tested on this passage after a 1-week interval, on Day 14.

## RESULTS AND DISCUSSION

### *Effect of Background Knowledge on Initial Parallel Learning*

As hypothesized, previously learned relevant background knowledge (Christianity) had a significant facilitating effect ($p < .05$) on the learning and retention of parallelly related Buddhism material (see Tables 1 and 2). Both experimental and control subjects in the upper part of the distribution of scores on the Christianity test made higher retention scores on the Buddhism test than did subjects in the lower part of the distribution of Christianity test scores. This facilitating effect of background knowledge was manifested when the influence of verbal ability was statistically controlled in a covariance design. The stability and clarity of related background knowledge therefore constituted a significant limiting condition with respect to learning the unfamiliar Buddhism material. It presumably did so by enhancing the discriminability of the new material. Similar findings were reported in a previous study (see Ausubel & Fitzgerald, 1961).

Another possible interpretation of this finding that cannot be excluded by our experimental design is that higher scores on both the Christianity and Buddhism tests are a common reflection of either (*a*) greater *general* knowledge about religious terms and concepts, which gives the generally more knowledgeable subjects an advantage in understanding, organizing, learning, and retaining both Christianity and Buddhism material; or (*b*) greater general interest in religious subject matter, which provides the more interested subjects with greater motivation to learn both bodies of religious knowledge. The discriminability hypothesis and these alternate explanations, however, are not mutually exclusive. Furthermore, the discriminability hypothesis is a more proximate interpretation of the finding than is the "general religious knowledge" hypothesis, because it involves the influence of more closely related and pertinent background material.

TABLE 1

MEAN RETENTION SCORES OF EXPERIMENTAL AND CONTROL GROUPS ON BUDDHISM TEST, ADJUSTED FOR SCAT SCORES AND ANALYZED BY CHRISTIANITY SCORE LEVELS

| Christianity score levels | Treatment group | |
|---|---|---|
| | Organizer | Control |
| Upper third | 20.6 | 18.7 |
| Middle third | 19.1 | 17.6 |
| Lower third | 18.5 | 16.5 |
| Total | 19.4 | 17.6 |

TABLE 2

COVARIANCE ANALYSIS OF BUDDHISM TEST SCORES × CHRISTIANITY SCORE LEVELS, ADJUSTING FOR SCAT SCORES

| Source | Adjusted | | | F |
|---|---|---|---|---|
| | SS′ | df′ | MS′ | |
| Treatment | 124.7 | 1 | 124.7 | 7.18** |
| Christianity score levels | 121.6 | 2 | 60.8 | 3.50* |
| Interaction | 5.3 | 2 | 2.6 | .15 |
| Between | 251.5 | 5 | 50.3 | 2.90* |
| Within | 2,691.7 | 155 | 17.4 | |

* $p < .05$.
** $p < .01$.

### *Effect of Initial Learning on Later Parallel Learning*

Consistent with the above finding regarding the effect of Christianity knowledge on Buddhism retention scores, knowledge of the newly learned Buddhism material significantly facilitated ($p < .01$) the learning and retention of the Zen Buddhism passage (see Tables 3 and 4). That is, with verbal ability statistically controlled, the subjects who made high, average, or low scores on the Buddhism test tended to make corresponding scores on the Zen Buddhism test. However, the facilitating effect of high stability and clarity of previously learned related knowledge on the discriminability of the new material was evidently greater in this instance than in the preceding instance because of the

TABLE 3

MEAN RETENTION SCORES OF EXPERIMENTAL AND CONTROL GROUPS ON ZEN BUDDHISM TEST, ADJUSTED FOR SCAT SCORES AND ANALYZED BY BUDDHISM SCORE LEVELS

| Buddhism score levels | Treatment group | |
|---|---|---|
| | Organizer | Control |
| Upper third | 17.4 | 16.9 |
| Middle third | 13.2 | 14.1 |
| Lower third | 13.1 | 11.6 |
| Total | 14.8 | 14.2 |

TABLE 5

MEAN RETENTION SCORES OF EXPERIMENTAL AND CONTROL GROUPS ON BUDDHISM TEST, ADJUSTED FOR CHRISTIANITY TEST SCORES AND ANALYZED BY SCAT LEVELS

| SCAT level | Treatment group | |
|---|---|---|
| | Organizer | Control |
| Upper third | 21.1 | 20.6 |
| Middle third | 18.7 | 16.5 |
| Lower third | 18.3 | 15.8 |
| Total | 19.4 | 17.6 |

greater similarity between the Buddhism and Zen Buddhism material than between the Christianity and Buddhism material. The same alternate interpretations considered above also apply to this finding.

*Effect of Organizer on Initial Learning*

Analysis of covariance indicates that the organizer treatment facilitated the learning and retention of the Buddhism passage. This effect was significant at the .01 level when verbal ability was held constant (see Table 2) and at the .05 level when Christianity scores were held constant (see Table 6). Contrary to our hypothesis and to previous findings (Ausubel & Fitzgerald, 1961), however, there was no significant interaction between Christianity scores and the organizer (see Table 2). The facilitating effect of the organizer on Buddhism retention scores was apparently uniform over the entire range of Christianity scores (see Table 1) instead of being differentially more beneficial, as in the previously cited study, for those subjects who, according to our hypothesis, should have found the new material less discriminable because of the relative instability and unclarity of their Christianity knowledge.

Table 6 also shows that the interaction between the organizer and verbal ability was not statistically significant. Nevertheless, as in a previous study (Ausubel & Fitzgerald, 1962), there was a suggestive trend for the organizer differentially to benefit the lower ability sub-

TABLE 4

COVARIANCE ANALYSIS OF ZEN BUDDHISM TEST SCORES × BUDDHISM SCORE LEVELS, ADJUSTING FOR SCAT SCORES

| Source | Adjusted | | | |
|---|---|---|---|---|
| | *SS'* | *df'* | *MS'* | *F* |
| Treatment | 21.5 | 1 | 21.5 | 1.41 |
| Buddhism score levels | 486.7 | 2 | 243.4 | 15.96* |
| Interaction | 16.0 | 2 | 8.0 | .52 |
| Between | 524.2 | 5 | 104.8 | 6.87* |
| Within | 2,363.4 | 155 | 15.3 | |

* $p < .01$.

TABLE 6

COVARIANCE ANALYSIS OF BUDDHISM TEST SCORES × SCAT LEVELS, ADJUSTING FOR CHRISTIANITY SCORES

| Source | Adjusted | | | |
|---|---|---|---|---|
| | *SS'* | *df'* | *MS'* | *F* |
| Treatment | 111.6 | 1 | 111.6 | 6.57* |
| SCAT levels | 474.1 | 2 | 237.1 | 14.04** |
| Interaction | 43.1 | 2 | 21.5 | 1.28 |
| Between | 628.8 | 5 | 125.8 | 7.45** |
| Within | 2,618.6 | 155 | 16.9 | |

* $p < .05$.
** $p < .01$.

jects; the only significant difference ($p < .05$) between organizer and control groups occurred in the lower-third subgroup of subjects on SCAT (see Table 5). Presumably the subjects of average and better ability are more capable of *spontaneously* discriminating between new ideas and previously learned related ideas, and hence derive less benefit from an advance organizer designed to enhance such discriminability.

### *Effect of Organizer on Later Leaning*

Unlike its effect on the Buddhism retention scores, the organizer did not significantly facilitate the learning and retention of the Zen Buddhism passage (see Table 4). This negative finding possibly reflected the fact that since the recently learned and tested Buddhism material was still relatively fresh in mind while the Zen Buddhism passage was being learned, the ideas in the two sets of material could be readily enough discriminated without the benefit of the organizer, especially by subjects with an average or better knowledge of Buddhism. Contrary to our hypothesis, also, there was no significant interaction between the organizer and level of Buddhism knowledge (see Table 4).

## REFERENCES

AUSUBEL, D. P., & FITZGERALD, D. The role of discriminability in meaningful verbal learning and retention. *J. educ. Psychol.*, 1961, **52**, 266–274.

AUSUBEL, D. P., & FITZGERALD, D. Organizer, general background and antecedent learning variables in sequential verbal learning. *J. educ. Psychol.*, 1962, **52**, 243–249.

(Received May 13, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 6, 337–345

# BYPASSING AS A WAY OF ADAPTING SELF-INSTRUCTION PROGRAMS TO INDIVIDUAL DIFFERENCES[1]

VINCENT N. CAMPBELL
*American Institute for Research, Palo Alto*

A self-instruction program on set theory was prepared in 3 forms. The bypass form allowed S to skip up to 85% of the learning material as a consequence of correct responses to 15 required basic steps. The long linear form permitted no bypassing and the short linear form presented only the basic steps. 780 Ss in 4th–12th grade classrooms were assigned 1 of the 3 learning methods and given a test of amount learned the next day. Test score and learning time means were highest for the long groups and lowest for the short groups. Amount learned and time spent corresponded closely to number of program pages studied. Bypassing was thus no more efficient than the linear methods, regardless of mathematical ability level.

Encountered casually, individual differences in personality and behavior are the spice of life. But to the educator or psychologist who confronts them on duty, they are a force to be reckoned with. In trying to cause the same change in many persons, or a difference between groups, he often finds that individual differences remain the largest source of variance in the outcome. Nowhere has this been more evident than in schoolroom learning. Ability testing, now almost universal in schools, is based on the premise that these individual differences will persist.

The present research is also based on this premise. Our purpose has been to ascertain whether programed self-instruction might be improved by varying the instructional sequence so as to adjust to individual differences among learners. Early pilot studies (Campbell, 1961) failed to demonstrate the practicality of employing parallel linear programs as a means for adjusting to individual differences, when variations in directness of cueing and mode of responding are the distinguishing features of parallel programs.

A reasonable alternative to parallel linear programs is some form of bypassing, that is, a branching program in which everyone is presented the minimum set of basic steps, and individuals who fail a given step immediately take a remedial "loop" of instruction on that step. One bypassing program may thus offer a large number of different sequences of instruction. In theory each learner receives a sequence which emphasizes his particular weak points and minimizes time spent in places where he excels through acumen or prior knowledge.

The appropriateness of bypassing for a given learning topic depends on the extent of individual differences in amount of instruction needed at each step. Given an appropriate topic, the success of bypassing should depend mainly on two things: the validity of each decision which directs the learner either to begin a remedial loop or to bypass it, and the effectiveness of the remedial instruction contained in the loops. In other words, one must accurately evaluate the learner's progress and also know what to do if his learning is inadequate.

---

[1] The research reported herein was conducted under a grant from the Office of Education, United States Department of Health, Education, and Welfare, L. J. Briggs being the principal investigator. We are greatly indebted to the following school districts for their cooperation and assistance with the research reported here: Hillsborough Elementary, Menlo Park Elementary, San Mateo Elementary, San Mateo Union High School, and Sequoia Union High School. Special thanks are due D. F. Terry, M. B. Willis, and J. W. Shearer for their help in program writing and research.

The principle of "adaptive" instruction is not unique to programing, of course. The same paradigm underlies all educational plans in which instruction is intended to be geared to the learner's progress. Aside from careful attention to sequencing of material and the fact that the instructor's help is not needed to adjust the procedure to the student's progress, the main feature distinguishing self-instruction programs is the greater frequency of the learner's responses through which his progress can be evaluated. The study of bypassing programs thus lends new perspective to the larger and older problem of remedial instruction in general. Remedial instruction is intended to be adaptive to the learner's progress, by definition. But incisive studies of its two main requirements, validity of evaluation and effectiveness of remedial learning, seem peculiarly rare in the literature (Curr & Gourlay, 1953). Perhaps this is because it seems obvious that extra instruction where the learner has failed will be beneficial. We set out to prove this truism in the realm of programed instruction and were taken aback by the findings reported here.

In earlier studies of branching programs (Campbell, 1961; Coulson & Silberman, 1960; Silberman, Coulson, Melaragno, & Estavan, 1960) bypassing and nonbypassing yielded equivalent criterion test score means, but in some of these experiments bypassing saved time and in others it did not. Because of the mixed findings of prior studies our main experiment was preceded by a series of pilot studies aimed at refining the experimental program and testing hypotheses about key characteristics of bypassing.

## DEVELOPMENT OF THE EXPERIMENTAL PROGRAM

The learning topic chosen was a part of set theory. The main objectives of the program were grasp of the fundamental concepts presented and ability to use these concepts to make translations in both directions between symbolic expressions and concrete situations. The criterion test items used to assess accomplishment of the learning objectives were heterogeneous in form and were not those used in the program itself.

In the initial program and successive revisions we tried to give as much care to writing the basic steps required of all learners as to the remedial loops. Finding that a well-written loop teaches more than a poorly written basic step would be of little interest. The main writing guide, beyond having the topic well in mind, was to keep it as clear and interesting as we could and in the simplest language possible. We tried to keep the basic steps concise in order to maximize the time that might be saved through bypassing. The remedial loops were changed and lengthened substantially during revision.

At no time did we try to conform to a particular school or style of programing. In view of the mixed and generally negative findings on the importance of particular features of programed instruction (e.g., Briggs, Goldbeck, Campbell, & Nichols, 1962; Lumsdaine & Glaser, 1960) we tried to fit the program to the particular learning objectives, altering the form pragmatically on the basis of several group experiments in classrooms (fourth to twelfth grades) and intensive tape recorded interviews with individual subjects. Aside from specific changes in content intended to make the instruction more effective, the major characteristics of the program which emerged as a result of these developmental studies and a review of related literature are listed below as principles:

1. Make the learner's response to each basic step a valid indicator of his learning progress, however difficult a response this may require.
2. Avoid failures early in the remedial loop (Valentine, 1951). Application of this principle took the form of requiring no responses at all, or only easy ones, in the first 2 or 3 pages of the loop.
3. Make a special point of discriminating and contrasting easily confused ideas and symbols, e.g., intersection ($\cap$) and union ($\cup$).
4. In teaching complex combinations, begin at an elementary level and take a slow part-by-part approach, only gradually combining the simpler ideas into a whole (Gagné, 1962).
5. Series of illustrations and examples are more effective for learning concepts and principles than are abstract or formal statements during initial learning (Haslerud & Meyers, 1958; Hendrix, 1946).

The final version of the sets program was prepared in a bypassing form and in two linear forms. The bypass form and the long linear form each consisted of 100 $7 \times 8\frac{1}{2}$ inch pages, of which 65 called for one or more responses. The correct answer always appeared on the back of the same page. The program was divided into 15 booklets corresponding to the 15 basic steps. The first page of any booklet was the basic step, which concluded by calling for a response. The remedial loop made up the rest of each booklet. The loop was required work in the long linear form, but was bypassable in the bypass form. The short linear form of the program consisted of the 15 basic step pages stapled together in a single booklet, the remedial loops being entirely excluded.

### *Criterion for Bypassing*

The format of the bypass program was the same as the long linear one except that on the back of every basic step page beside the correct answer was printed this instruction: "If your answer was *right*, skip to the next booklet *now*." Where two or more answers were called for, the subject was instructed to skip only if all were correct. Midway through four of the longest loops the subject was given a second chance to bypass the rest of the loop in this manner.

Besides correctness of the learner's response, another potentially important criterion for bypassing is the learner's own judgment as to how well he has learned. We compared these modes of evaluation with the sets program in three eighth grade classes. In addition to a group bypassing on the basis of response correctness, another group bypassed each remedial loop only if they both answered correctly and "understood the page." Since program responses were apparently easier than posttest items, and since "understanding" might be subjective enough to be swayed by motivation, we thought it more promising to require both understanding and response correctness than to make understanding alone sufficient for bypassing.

There was no difference between the two groups in mean test score, nor in mean learning time. The new bypassing criterion had no apparent advantage and had the disadvantage of being more complicated, so we used only response correctness in the main experiment.

### *Effect of Evaluation on Remedial Learning*

A common observation is that initial incomprehension of something mathematical often induces panic, indifference, hostility, or some other unreceptive state of mind, which in many people becomes a permanent attitude toward mathematics. Since each new idea in the sets program was introduced through a concise basic step and a question difficult enough to be a valid index of learning, it seemed worthwhile to investigate the possibility that failing the basic step might hinder subsequent learning from the remedial loop.

In order to test this hypothesis we prepared "reversed" long programs in which each remedial loop was studied before its basic step rather than after it. We then compared groups who took this reversed long linear method with groups who took the long linear program in its normal order (basic step, then loop). Both methods presented the same learning materials, but the reversed long method led the student gradually from the simple to the difficult, hopefully thereby avoiding adverse effects of immediate failure on that segment. The two methods were compared in two fourth grade and two eighth grade classes. There were no appreciable differences between the two methods in test score or learning time at either grade level. Apparently the basic step did not detract from remedial learning with the sets program.

Details of the studies on effects of evaluation, bypassing criteria, and program development in general are reported elsewhere (Campbell, 1962).

## METHOD

### *Subjects*

A total of 780 public school students from the San Francisco Peninsula area served as subjects for the main experiment on bypassing. The classes chosen were essentially representative samples of students from these particular schools and grade levels as nearly as we could judge by administrative criteria and ability scores. The breakdown by grade level and school was as follows: the entire fourth grade ($n = 166$) and sixth grade ($n = 173$) populations of two schools, 150 eighth graders from a third school, and, from one high school, 150 tenth graders and 139 twelfth graders.

### *Design*

The effectiveness of bypassing with the sets program was judged by comparing criterion test scores and time spent learning for the three learning methods (program forms), which are summarized below:

1. Bypass: After each basic step the subject skipped to the next basic step if his answer was right. If his answer was wrong he continued on through the remedial loop of instruction.

2. Short (linear): The program contained no remedial loops. The subject worked through the 15 basic steps in linear fashion.

3. Long (linear): The subject worked straight through the entire program of basic steps and remedial loops with no opportunity to skip any of the loops.

### *Procedure*

In each fourth, sixth, and eighth grade class the subjects were assigned either the bypass, long, or short form of the program by passing out the three forms in alternation. In the same manner the tenth and twelfth grade subjects were assigned either the bypass program or the short program. The long linear method was omitted for the high school samples because of the difficulty of rescheduling classes to permit sufficient time.

The instructions on the front page of every program briefly explained the format of the sets lesson and emphasized the educational (as opposed to testing) purpose of the "exercises" therein. Bypassing was explained (to bypass subjects only) as a way "to spend the most time on the ideas that you find hardest."

All subjects in a class began their programs at the same time and each worked at his own pace until he had finished, except for the few (about 8%) who did not quite finish in the maximum time available. The latter subjects were included in all data analyses. Normal recesses were taken in the fourth, sixth, and eighth grades. Time spent on the program was recorded for each subject. One third of the twelfth grade sample was kept especially quiet throughout learning to assess the effects of noise on learning. It had been observed that noise increases toward the end of a class period, and that this might differentially affect the learning methods.

The sets criterion test was given the following day at the same hour. Nearly all subjects spent between 10 and 20 minutes on the test, but there was no time limit.

## Results

At all grade levels bypass group criterion test means surpassed short group means and were lower than long group means (Grades 4, 6, and 8). Likewise, mean learning times for bypass groups fell between those of the short and long linear groups. Figure 1 shows mean test score as a function of mean learning time for each method at each grade level. At any given grade level the points for the short and long methods may be thought of as two points on a curve representing amount learned as a function of length of a linear program covering a fixed set of objectives. It is difficult to estimate exactly where the interpolated point would fall on such a curve for a linear program of intermediate length requiring about the same time to complete as the bypass program. But judging by the general shape of learning curves of this type, a medium length linear program would teach at least as much as the bypass program if not more.

In order to detect ability effects, each grade in the sample was divided into three approximately equal sized ability groups on the basis of available mathematical ability scores. The California Achievement Test (CAT) Total Arithmetic score was used in the fourth and sixth grades; and the Differential Aptitude Test (DAT) Numerical score in the eighth, tenth, and twelfth grades. Mean criterion test scores and learning times for each method at each grade and ability level are shown in Tables 1 and 2. Test score and learning time variances

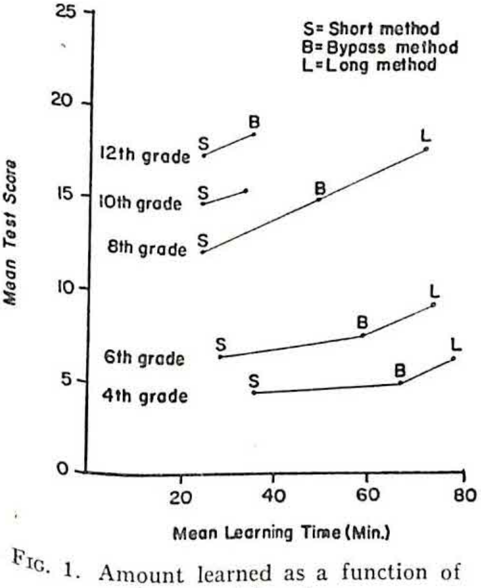


Fig. 1. Amount learned as a function of time spent learning.

TABLE 1

Sets Test Means for Each Learning Method, Grade, and Ability Level

| Grade | Ability | Short | Bypass | Long |
|---|---|---|---|---|
| | | Learning method | | |
| Twelfth | | | | |
| | High | 27.1 | 26.3 | |
| | Medium | 17.9 | 16.9 | |
| | Low | 11.8 | 12.9 | |
| Tenth | | | | |
| | High | 20.6 | 23.7 | |
| | Medium | 11.5 | 12.1 | |
| | Low | 10.0 | 11.1 | |
| Eighth | | | | |
| | High | 17.9 | 20.7 | 24.3 |
| | Medium | 10.3 | 14.6 | 15.7 |
| | Low | 7.5 | 11.8 | 11.5 |
| Sixth | | | | |
| | High | 9.4 | 10.1 | 12.0 |
| | Medium | 5.4 | 7.6 | 9.5 |
| | Low | 4.3 | 3.8 | 6.0 |
| Fourth | | | | |
| | High | 7.8 | 6.5 | 7.9 |
| | Medium | 4.7 | 5.3 | 6.7 |
| | Low | 2.1 | 1.9 | 3.2 |

Note.—*n* per entry varies from 12 to 28. Estimated *SD* within methods is approximately 5.0.

TABLE 2

LEARNING TIME MEANS (IN MINUTES) AND STANDARD DEVIATIONS FOR EACH LEARNING METHOD, GRADE, AND ABILITY LEVEL

| Grade | Ability | Short *M* | Short *SD* | Bypass *M* | Bypass *SD* | Long *M* | Long *SD* |
|---|---|---|---|---|---|---|---|
| Twelfth | | | | | | | |
| | High | 23 | 9 | 30 | 10 | | |
| | Medium | 23 | 5 | 33 | 8 | | |
| | Low | 26 | 5 | 39 | 7 | | |
| Tenth | | | | | | | |
| | High | 23 | 5 | 29 | 8 | | |
| | Medium | 25 | 8 | 36 | 8 | | |
| | Low | 26 | 5 | 34 | 9 | | |
| Eighth | | | | | | | |
| | High | 24 | 3 | 38 | 8 | 66 | 11 |
| | Medium | 25 | 5 | 46 | 13 | 76 | 15 |
| | Low | 25 | 5 | 57 | 15 | 71 | 19 |
| Sixth | | | | | | | |
| | High | 27 | 9 | 60 | 17 | 82 | 13 |
| | Medium | 27 | 5 | 60 | 16 | 68 | 16 |
| | Low | 30 | 10 | 58 | 17 | 77 | 14 |
| Fourth | | | | | | | |
| | High | 35 | 9 | 62 | 23 | 78 | 22 |
| | Medium | 35 | 8 | 67 | 28 | 79 | 22 |
| | Low | 38 | 11 | 74 | 23 | 77 | 20 |

were then analyzed with ability group as an independent variable, which required correcting for disproportionate *N*s. Since only two learning methods were used in the high school, two overlapping analyses of variance were performed for each dependent variable. Both analyses included all three ability groups, but one analysis compared the short and bypass methods across all five grade levels while the second analysis compared all three learning methods in the fourth, sixth, and eighth grades. The summaries of these analyses are shown in Table 3 for sets test scores and Table 4 for learning times.

These analyses gave no indication at all that the effectiveness of bypassing relative to the linear methods changes as a function of the student's level of mathematical ability. There was a slight tendency ($p < .05$) for high ability students to finish the program sooner, but this difference was almost entirely accounted for by the bypass method alone. It makes sense that the more able students would tend to make fewer errors

TABLE 3

ANALYSES OF VARIANCE OF SETS CRITERION TEST SCORES

| Source | Short versus bypass (all grades) *df* | *MS* | *F* | All learning methods (fourth, sixth, and eighth grades) *df* | *MS* | *F* |
|---|---|---|---|---|---|---|
| Grade (G) | 4 | 3630 | 110.5*** | 2 | 3714 | 166.6*** |
| Ability (A) | 2 | 3998 | 121.7*** | 2 | 1858 | 83.4*** |
| Method (M) | 1 | 175 | 5.3* | 2 | 373 | 16.8*** |
| G × A | 8 | 147 | 4.5*** | 4 | 156 | 7.0*** |
| G × M | 4 | 77 | 2.4 | 4 | 83 | 3.7** |
| A × M | 2 | 0 | — | 4 | 0 | — |
| G × A × M | 8 | 10 | — | 8 | 3 | — |
| Within | 489 | 33 | | 427 | 22 | |
| Total | 518 | | | 453 | | |

* $p < .05$.
** $p < .01$.
*** $p < .001$.

TABLE 4

ANALYSES OF VARIANCE OF LEARNING TIMES

| | Short versus bypass (all grades) | | | All learning methods (fourth, sixth, and eighth grades) | | |
|---|---|---|---|---|---|---|
| Source | *df* | *MS* | *F* | *df* | *MS* | *F* |
| Grade (G) | 4 | 10728 | 68.4*** | 2 | 6873 | 26.5*** |
| Ability (A) | 2 | 1396 | 8.9*** | 2 | 257 | — |
| Method (M) | 1 | 57036 | 363.5*** | 2 | 78301 | 301.4*** |
| G × A | 8 | 91 | — | 4 | 480 | 1.9 |
| G × M | 4 | 2913 | 18.6*** | 4 | 530 | 2.0 |
| A × M | 2 | 609 | 3.9* | 4 | 474 | 1.8 |
| G × A × M | 8 | 147 | — | 8 | 339 | 1.3 |
| Within | 489 | 157 | | 427 | 260 | |
| Total | 518 | | | 453 | | |

Note.—Within-group variances were heterogeneous but no useful transformation of learning times was apparent. This may add an increment of doubt to marginally significant effects such as the A × M interactions.

* $p < .05$.

*** $p < .001$.

and thus bypass oftener and finish sooner. Bypass subjects at higher grade levels should also have finished sooner for this reason, which appears to account for most of the Grade × Method interaction indicated in Table 4. Although relative learning efficiency of the bypass, long, and short methods did not depend appreciably on ability level, there were substantial differences between ability levels in amount learned for all methods. There were also marked differences between grade levels in amount learned, but learning time differences between grades were very slight for the linear methods from the sixth grade up.

Comparison of sets test means for the "noisy" versus "quiet" rooms in the twelfth grade sample showed no sign at all of greater learning in the quiet room, either for the group as a whole or for the bypass group relative to the short group. Apparently social distraction was not a notable factor in determining amount learned. This result corroborated those of earlier pilot studies in which we tried to avoid social distraction by giving the subjects who finished early another task to keep them occupied.

## DISCUSSION

As a way of adapting self-instruction to individual differences in order to increase learning efficiency, bypassing did not manifestly succeed with the sets program. Even in the eighth grade where the remedial loops taught the most (see Figure 1), learning was no more efficient when the loops were required only after errors (bypass) than when they were required after all basic steps (long). Number of program pages studied, or average study time required, seemed to account for most of the variance in amount learned among the bypass, short linear, and long linear methods. An interesting question unanswered by the present research is whether studying the remedial loops would have been any more effective than repetition of the basic steps for an equivalent amount of time.

Since the remedial loops were of some benefit regardless of how they were assigned, it is appropriate to ask whether the fault lay in the basic step evaluations of learning progress. It is difficult to determine separately the validity of each evaluational response in the program, because there was not a one-to-one relationship between program items and test items. But since the program as a whole was meant to cover the same objectives as the test, the correlation between total errors on the short linear program and total test errors is a gross index of the validity of the evaluational responses as indexes of learning progress. The program-test correlations were in the vicinity of .50 for the eighth, tenth, and twelfth grades, indicating a substantial degree of validity.

Apart from the fact that program questions and test items are small "independent" samples of a very large population of behaviors indicating knowledge about sets, there are other sources of attenuation in these validity coefficients. The correlation between total test score and total program score glosses over differences among subjects as to which particular steps were failed, however validly such differences are diagnosed. Another source of attenuation is the effect of feedback (answers) following the basic step responses.

The program-test correlations were so small as to be negligible for the fourth and sixth grades. The younger students may not have understood or accepted the bypassing procedure as easily as older students. For example, about 20% of all bypassing decisions were in error (e.g., bypassing after a wrong response) and a disproportionate number of these occurred at the lower grade levels. This may also explain why the fourth and sixth grade bypass groups learned even less efficiently than might have been expected of a linear group taking the same time.

There were fewer wrong answers proportionally on the program than on the criterion test, which suggests another factor that may have limited the validity of the program score and the usefulness of bypassing. Perhaps the basic steps were demanding enough to require momentary grasp of the idea, but did not require enough practice and thinking through to insure 24-hour retention and transfer to new problems, especially in an area as unfamiliar as set theory was to most of these subjects. If this were the case, bypassing might be considerably more useful as a review technique to be used after initial instruction has created a wider span of individual differences in knowledge. Similarly, increasing severalfold the amount of basic and remedial instruction contained in each of the 15 steps of the sets program might have enhanced the efficiency of bypassing.

Unfamiliarity with the bypassing technique, as well as the content, may have worked against bypassing in the present studies. A more longitudinal investigation of the use of bypassing is needed to determine whether practice in the use of bypassing would make it more valuable.

One reason for choosing set theory as an experimental topic was that the steps are somewhat sequentially dependent. It seemed that the on-the-spot remedial self-instruction provided by bypassing self-instruction programs would be especially beneficial for such a hierarchical topic. There may be other relevant factors which would enable one to select learning topics and objectives more amenable to bypassing in initial instruction than was set theory.

One such variable may be the extent to which different learners fail different steps within a topic. We developed a statistical index (Campbell, 1962) similar to Guttman's coefficient of reproducibility (Green, 1954) which quantitatively expresses this interdependence of difficulty of various parts of a learning topic for a given sample of learners. We applied the index to data from an early version of the sets program and a program on the United Nations. The results indicated greater individual differences in relative difficulty of the steps in the United Nations program, suggesting a possible advantage over set theory in the use of bypassing.

On this basis we conducted preliminary studies (Campbell, 1962) comparing bypassing and short linear forms of the United Nations program. In eighth grade and twelfth grade classes bypass groups learned more than short linear groups but took considerably more time in learning. The results resembled those with the sets program in the present high school sample. In both cases further research is needed to determine whether bypassing is more efficient than giving all subjects the remedial loops in linear format.

If bypassing has any value for any topic, either for initial learning or review, it has yet to be shown. Perhaps the generally accepted rationale for adaptive and remedial instruction deserves critical reappraisal. In any case the problem cannot practically be ignored. There are core subject matters in which we want all students to learn the fundamentals. Branching and bypassing in self-instruction programs may not prove to be effective for any given one, but the problem of how those who fail initially can learn to a minimum criterion of mastery will remain.

## REFERENCES

Briggs, L. J., Goldbeck, R. A., Campbell, V. N., & Nichols, D. G. Experimental results concerning form of response, size of steps, and individual differences in automated programs. In, *Proceedings of the Symposium on the Application of Digital Computers to Automated Instruction.* New York: Wiley, 1962.

Campbell, V. N. Adjusting self-instruction programs to individual differences: Studies of cueing, responding and bypassing. (Publ. No. AIR-C41-7/61-SR) San Mateo: American Institute for Research, 1961.

Campbell, V. N. Studies of bypassing as a way of adapting self-instruction programs to individual differences. (Publ. No. AIR-C41-5/62-FR) San Mateo: American Institute for Research, 1962.

Coulson, J. E., & Silberman, H. F. Results of an initial experiment in automated teaching. *J. educ. Psychol.,* 1960, **51**, 135–143.

Curr, W., & Gourlay, N. An experimental evaluation of remedial education. *Brit. J. educ. Psychol.,* 1953, **23**, 45–57.

Gagné, R. M. The acquisition of knowledge. *Psychol. Rev.,* 1962, **69**, 355–365.

Green, B. Attitude measurement. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge: Addison-Wesley, 1954.

Haslerud, G. M., & Meyers, S. The transfer value of given and individually derived principles. *J. educ. Psychol.,* 1958, **49**, 293–298.

Hendrix, G. A new clue to transfer of training. *Elem. Sch. J.,* 1946, **48**, 197–208.

Lumsdaine, A. A., & Glaser, R. (Eds.) *Teaching machines and programmed learning.* Washington, D. C.: National Education Association, 1960.

Silberman, H. F., Coulson, J. E., Melaragno, R. J., & Estavan, D. P. Fixed sequence versus branching in a computer-based teaching machine. (Report No. SP-195) Santa Monica: System Development Corporation, 1960.

Valentine, H. B. Some results of remedial education in a child guidance centre. *Brit. J. educ. Psychol.,* 1951, **21**, 145–149.

(Received May 13, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 6, 346–355

# STUDENT TEACHER CHARACTERISTICS FROM THE PUPILS' VIEWPOINT[1]

DONALD J. VELDMAN AND ROBERT F. PECK
*University of Texas*

Five factors were extracted from evaluations of 554 student teachers on the 38-item Pupil Observation Survey: I. Friendly, Cheerful, Admired; II. Poised, Knowledgeable; III. Interesting, Preferred; IV. Strict Control; and V. Democratic Procedure. Invariance of factor structure was demonstrated across 3 semester subsamples and across teacher-sex, and factor-score reliabilities were computed. Female teachers were higher than males on Dimensions I and V. Correlations with 2 attitude inventories were low, but significant. Relationships of factor-scores to supervisor-rated effectiveness were linear and positive for Factors I, II, and IV, except for a curvilinear effect with males on Factor I.

The description and evaluation of teacher behavior continues to be a central problem in educational research (Barr, Worcester, Abell, Beecher, Jensen, Peronto, Ringness, & Schmid, 1961), and a recent book by Sarason, Davidson, and Blatt (1962) focuses particular attention on the evaluation of teacher preparation programs. Many schemes have been devised for converting subjective impressions derived from classroom observation into quantified statements. The present study describes the development of an instrument designed to reduce the everyday observations of pupils to reliable measures of important aspects of teacher behavior.

One major difficulty often encountered in assessment operations of this kind has been the choice of dimensions to represent the most relevant aspects of teaching performance. Some investigators have resorted to a simple summation of ratings on separate scales in an attempt to obtain a stable criterion measure (Dalton, 1962). Although a strong general factor usually appears among scales which attempt to measure aspects of teaching effectiveness, it is sometimes possible to extract more than one meaningful dimension from a series of such judgements. Cluster and factor analytic methods have been employed occasionally for this purpose. For instance, Ryans' (1960) very thorough research monograph describes three major factors of teacher-classroom behavior which emerged from an extensive investigation of ratings by trained observers previously experienced in teaching:

Pattern Xo: understanding, friendly versus aloof, egocentric, restricted teacher behavior

Pattern Yo: responsible, businesslike, systematic versus evading, unplanned, slipshod teacher behavior

Pattern Zo: stimulating, imaginative, surgent, or enthusiastic versus dull, routine teacher behavior

In contrast to Ryans' observers, the pupils in the classroom are untrained as raters and, of course, have no teaching experience. One of the weaknesses inherent in most studies involving the use of trained observers is the limited sample of actually observed behaviors upon which the assessments must be based. Adequate evaluation demands an adequate sample of observed performance in the wide variety of roles demanded of the modern teacher. A single hour of classroom observation seldom provides sufficient evidence for an observer to evaluate the full range of attributes of interest to investigators.

---

[1] This research is a part of the Mental Health in Teacher Education Project, College of Education, University of Texas, supported by Grant No. 2M-6635 from the Training Branch, National Institute of Mental Health. Portions of this paper were presented at the meetings of the American Educational Research Association, Chicago, February 1963.

No one would seriously argue that the pupils taught by a teacher are inevitably the best judges of her effectiveness or ability. Nevertheless, the pupils have one major advantage over other observers: they see the teacher perform on many different occasions, as she encounters a wide variety of problems, as she attempts quite varied tasks, and as she deals with individuals known personally to the observer. Not only does each pupil have the advantage of many separate observations upon which to base his judgement, the use of pupils as observers also affords the increased reliability and reduction of bias that multiple judges afford. The validity of pupil observations remains to be demonstrated, but there are strong a priori reasons to expect higher reliability from this method than from procedures employing single observers.

In the assessment program of the Mental Health in Teacher Education project at the University of Texas a number of different observation methods have been used in attempting to gather comprehensive pictures of student teacher behavior. One of the techniques developed by the project staff is a questionnaire called the Pupil Observation Survey (POSR). During three semesters this form was completed anonymously by each of the pupils taught by each of the student teachers in the junior and senior high schools of Austin, Texas. The present paper concerns the dimensions of student teacher behavior which emerged from analyses of these data. Item means for each teacher were factor analyzed, and factor scores were computed for each student teacher. The nature and stability of the factor dimensions, their reliability as scaled scores, and preliminary validity evidence will be reported.

## METHOD

The POSR is a 38-item questionnaire presenting a 4-point scale of agreement after each statement. Many of the items were adapted from an earlier experimental form (McClain, 1961; McClain & Bown, 1961). All statements are phrased positively and attempt to measure pupils' reactions, as well as behavioral characteristics of the student teacher. Separate forms, differing only in the relevant pronouns, were prepared for male and female student teachers. The instructions on the face sheet of the form were as follows:

> To The Student:
>
> The College of Education at the University of Texas is studying several ways of preparing student teachers. The University wants to know how you have felt about working with your student teacher this semester. Your answers to these questions will be used for research only, and will have nothing to do with the grade your student teacher gets. When you answer the questions on the next pages, it is important to mark the way you really feel.
>
> After you read each question, put a circle around *one* of the four letters that follow each statement. If you think the statement is completely true, circle the capital T. If you think the statement is more true than it is false, put a circle around the small t. If you think the statement is more false than it is true, circle the small f. If you think the statement is completely false, circle the capital F.
>
> Here is an example:
>
> My student teacher has a loud voice. T t f F
>
> If you think she has a very loud voice, you would circle the big T: Ⓣ t f F
>
> If her voice is a little louder than most people, circle the small t: T ⓣ f F
>
> If her voice is a little softer than most people, circle the small f: T t ⓕ F
>
> If her voice is very soft, you would circle the big F: T t f Ⓕ

When you finish the second page of questions, please go back and make sure that you did not accidentally skip any of them. Please answer all of the questions.

Thank you

The forms were administered to three semester subsamples, a total of 554 student teachers, by their public school cooperating teachers after the students had completed their required teaching sequences. Only classes from the seventh through twelfth grades were used. Considerable effort was expended to communicate to the pupils the fact that the evaluations would be utilized solely for research purposes, and very few cooperating teachers or student teachers refused to allow the forms to be administered.

A series of studies was carried out with these data in order to establish the usefulness of this method of obtaining descriptions of student teacher performance.

1. Factor analysis of POSR items. The circled choices on each pupil's protocol were punched on cards as scores on a 4-point scale. These item scores were reduced to item means for each student teacher through the use of a computer program which excluded unanswered items in the computation of the 38 means. The items were treated as "tests" and subjected to factor analysis,[2] using all student teacher means as data.

2. Factorial invariance. The three semester subsamples were independently factor analyzed in order to obtain evidence regarding invariance of factor structure over subject populations. All male teachers and all female teachers were also utilized in separate factor analyses. A factor matching procedure (Ahmavaara, 1954) was used to compare the factor structures from the three semester subsamples, as well as those derived from male and female teachers.

3. Factor score reliability. Factor scores were computed for each student teacher, on each dimension as it emerged in the factor analysis of POSR items with the total sample. The 50 teachers with the largest classes were selected, and item means and factor scores were recomputed separately for each randomly divided half of each teacher's class. These factor scores were then correlated to yield reliability estimates for the factor scores.

4. Sex differences on factor dimensions. The factor scores for male and female teachers were compared by single classification analysis of variance on each factor dimension.

5. Correlations with attitude inventories. The factor dimensions were used as criteria in computing zero-order and multiple correlations from the 18 scales of the California Psychological Inventory (CPI; Gough, 1957) and 10 scores from the Self-Report Inventory (SRI; Bown, 1961).

6. Validation against supervisor ratings of effectiveness. Ratings of the effectiveness of the student teachers by their University supervising professors were used as a criterion for each factor dimension. Because of strong central tendencies among the ratings on the various scales used, and because of extreme variations in central tendency among supervisors in their use of the rating form, the ratings on individual scales of the 5-item form were averaged (15 items for the first semester subsample). These means were then divided into approximately equal groups of student teachers for each supervisor: high, middle, and low rated.

## RESULTS

### *Factor Analysis of POSR Items*

Table 1 contains the factor structure of the 38 POSR items from the first analysis of all 554 student teachers.

The first factor was tentatively labeled Friendly, Cheerful, Admired and seems to be very much like Ryans' Pattern Xo, at a descriptive level. The second factor was called Knowledgeable, Poised and seems to describe self-confidence based on thorough understanding of the material to be taught. The third factor was labeled Interesting, Preferred and appears to match Ryans' Pattern Zo. The item content reflects lively, skillful presentation of materials as well as a generalized preference for teachers with these qualities. The fourth factor, called Strict Control, seems to resemble Ryans' Pattern Yo more than any of the other factors. These items could signify domineering behavior, but could also reflect a serious well-organized approach to teaching. The fifth factor, tentatively

---

[2] Principal-axis factor analysis, utilizing unities in the diagonal of the intercorrelation matrix, was used throughout this study. All factors with eigen values greater than unity were selected for rotation by the normalized varimax technique.

TABLE 1

FACTOR STRUCTURE OF POSR ITEMS FOR ALL 554 STUDENT TEACHERS

| Item | Loading |
|---|---|
| I. Friendly, Cheerful, Admired | |
| 26. She is always friendly toward her students | .90 |
| 38. She is good-natured and easy to get along with | .87 |
| 18. She never seems to order her students around | .85 |
| 29. She always seems cheerful and happy | .85 |
| 9. She usually looks on the bright side of things | .84 |
| 19. She smiles most of the time | .81 |
| 16. You can tell that she really likes her students | .78 |
| 37. She is always interested in hearing a student's ideas | .75 |
| 13. You can depend on her to be fair with you | .74 |
| 35. She is as interested in her students as she is in her subject | .74 |
| 6. She seems to understand the problems students have | .74 |
| 30. I would like to have her as a personal friend | .73 |
| 1. She is admired by most of her students | .67 |
| 21. She sets a good example for her students | .65 |
| 5. She hardly ever gets flustered about anything that happens | .64 |
| 11. I would like to be like her in some ways | .57 |
| II. Knowledgeable, Poised | |
| 17. She knows a great deal about her subject | .85 |
| 34. She doesn't get confused by unexpected questions | .77 |
| 27. She must have studied hard to know so much about her subject | .73 |
| 7. She is never stumped by a student's question | .72 |
| 36. She seems to know more about her subject than just what is in the book | .69 |
| 15. She always seems sure of herself in front of the class | .69 |
| 33. She always seems to know just what she'll do next | .63 |
| 25. She doesn't seem to be afraid of making mistakes | .53 |
| III. Interesting, Preferred | |
| 31. She makes learning seem more like fun than work | .76 |
| 10. She is the best teacher I have ever had | .73 |
| 12. Her class is never dull or boring | .72 |
| 20. I wish all my teachers were like her | .71 |
| 22. She knows how to put her subject across in a lively way | .70 |
| 2. She has made her subject alive and interesting for me | .70 |
| 24. She doesn't try to cover the lesson too fast | .61 |
| 4. She explains her assignments clearly and completely | .57 |
| IV. Strict Control | |
| 32. She doesn't let her students get away with anything | .86 |
| 3. She expects a lot from her students and usually gets it | .75 |
| 23. Students respect her because she means what she says | .66 |
| 14. She doesn't let the class discussion get too far off the subject | .57 |
| V. Democratic Procedure | |
| 8. Before she decides on a new project, she often asks the students what they think | .74 |
| 28. She likes to give the students a choice of how to do an assignment | .64 |

labeled Democratic Procedure, is loaded heavily by only two items, which have in common the teacher's active solicitation of the help of the pupils in making decisions about the goals and procedures to be chosen.

### *Factorial Invariance*

Table 2 presents the results of the factor matching analyses of the three semester subsamples. Since the invariance was so strong across these factor structures, only the coefficients for corresponding factors are shown in Table 2. The results of the factor matching of male and female student teachers' factor structures are shown in Table 3.

Factors I, II, and IV in both analyses appear to be highly similar. The two large coefficients for Female Factor III appear to be the result of a splitting of this factor into two components in the males' analysis. The items thus separated are those appearing in Table 1 as general Factor V (Democratic Procedure), an independent dimension which did not emerge in the females' analysis. For female teachers the two items reflecting willingness to allow pupils to take part in decisions about goals and methods appear as part of the dimension registering the pupils' liking for the teachers who are lively and interesting. This aspect of teacher behavior constitutes an independent dimension for male teachers, however, and is by definition *not* systematically related to the third factor. Factor VI in the males' analysis appears to be the result of a single item (14: "He doesn't let the class discussion get too far off the subject") which splits off from Female Factor IV. Why this item should be a symptom of strict class control for one sex and not the other is not apparent. Except for the minor variations noted above, the factor structures obtained separately for the two sexes appear to be almost identical.

TABLE 2

FACTORIAL EQUIVALENCE AMONG THREE SEMESTER SUBSAMPLES (S)

| Factor | $S_1$ versus $S_2$ | $S_1$ versus $S_3$ | $S_2$ versus $S_3$ |
|---|---|---|---|
| I | .99 | .99 | .98 |
| II | .99 | .99 | .98 |
| III | .96 | .98 | .98 |
| IV | .99 | .98 | .97 |
| V | .89 | .92 | .99 |

TABLE 3

FACTOR MATCHING COEFFICIENTS FOR COMPARISON OF FACTOR STRUCTURES FOR MALE ($N = 111$) AND FEMALE ($N = 443$) STUDENT TEACHERS

| | Females | | | |
|---|---|---|---|---|
| Males | I | II | III | IV |
| I | .99 | .06 | .03 | .12 |
| II | .07 | .98 | .20 | .02 |
| III | .14 | .33 | .90 | .25 |
| IV | .01 | .05 | .10 | .99 |
| V | .06 | .30 | .93 | .19 |
| VI | .18 | .33 | .47 | .80 |

### *Factor Score Reliability*

The "split-class" reliability coefficients calculated for each of the factor scores were: I. .92, II. .72, III. .91, IV. .81, and V. .89. These coefficients were not corrected with the Spearman-Brown formula for half-length tests since only the largest classes were used in obtaining these coefficients. These reliabilities appear high enough to warrant the use of the factor scores in subsequent validity studies.

### *Sex Differences on Factor Dimensions*

Table 4 contains the results of analyses of variance comparing the factor scores of male and female

TABLE 4

TEACHER-SEX DIFFERENCES IN POSR FACTOR SCORES

| Factor | Male $M$ | Female $M$ | $F$ |
|---|---|---|---|
| I. Friendly, Cheerful, Admired | 47.05 | 50.74 | 12.28** |
| II. Knowledgeable, Poised | 49.75 | 50.06 | .08 |
| III. Interesting, Preferred | 51.15 | 49.71 | 1.83 |
| IV. Strict Control | 49.44 | 50.14 | .43 |
| V. Democratic Procedure | 48.25 | 50.44 | 4.28* |
| $N$ | 111 | 443 | |

* $p < .05$.
** $p < .01$.

teachers. The highly significant difference for the first factor (Friendly, Cheerful, Admired) and the significant difference for the fifth factor appear to be consistent with role expectations in our culture; women are commonly considered to be more warm, exuberant, and permissive in their relationships with children.

### *Correlations with Attitude Inventories*

Futher information regarding the nature of POSR factors is provided by the correlations of CPI and SRI scales with each of the factor dimensions, separately for the male and female student teachers. These coefficients, as well as the multiple correlations and corresponding $F$ ratios, are located in Table 5 and 6.

TABLE 5

ZERO-ORDER AND MULTIPLE CORRELATIONS OF 18 CPI SCALES WITH EACH POSR FACTOR

| CPI scale | Males | | | | | Females | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | I | II | III | IV | V | I | II | III | IV | V |
| Do | −28** | 09 | 07 | 02 | 12 | 11* | 13** | 08 | 04 | 08 |
| Cs | −15 | 21* | −07 | −01 | 14 | 11* | 09* | 03 | 01 | 08 |
| Sy | −10 | 09 | 10 | −02 | 06 | 14** | 03 | 10* | 04 | 04 |
| Sp | −02 | −04 | 00 | −06 | 12 | 11* | 09* | 11* | 02 | 06 |
| Sa | −13 | 16 | 05 | −03 | 18 | 07 | 06 | 14** | 02 | 07 |
| Wb | −12 | 00 | 07 | −07 | −08 | 08 | −03 | −01 | −04 | −02 |
| Re | −11 | 22* | −11 | −16 | 00 | 05 | 06 | −11* | 00 | 12** |
| So | −04 | 08 | 07 | 03 | −15 | 08 | −01 | 05 | −07 | −01 |
| Sc | −01 | −01 | −13 | −01 | −15 | 04 | −06 | −07 | −06 | 00 |
| To | −08 | 07 | −02 | −04 | −06 | 10* | 02 | −01 | −05 | 05 |
| Gi | −03 | −03 | −05 | 03 | −02 | 09* | 00 | 00 | −09* | 04 |
| Cm | 00 | 08 | 04 | −09 | 01 | −03 | −03 | 08 | −02 | −05 |
| Ac | −18 | 08 | −03 | −01 | −04 | 07 | 10* | 04 | 03 | 02 |
| Ai | −09 | 13 | −29** | −28** | 11 | 02 | −01 | −15** | −03 | 08 |
| Ie | −08 | 12 | −18 | −27** | 10 | 11* | 04 | −04 | −02 | 02 |
| Py | −14 | −07 | −20* | −25** | 04 | −14** | −03 | −10* | −01 | 06 |
| Fx | 16 | −14 | −19* | −24* | 13 | 07 | −11* | −12** | 03 | 10* |
| Fe | 01 | 17 | −13 | −11 | 05 | 03 | −06 | −05 | −09* | 06 |
| $R$ | .399 | .398 | .494 | .532 | .437 | .315 | .304 | .305 | .217 | .227 |
| $F$ | .92 | .91 | 1.56 | 1.90* | 1.14 | 2.48** | 2.28** | 2.31** | 1.11 | 1.23 |

Note.—Males $N = 106$; females $N = 424$.
* $p < .05$.
** $p < .01$.

TABLE 6

ZERO-ORDER AND MULTIPLE CORRELATIONS OF 10 SRI SCORES WITH EACH OF THE POSR FACTORS

| | Males | | | | | Females | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SRI score | I | II | III | IV | V | I | II | III | IV | V |
| Self | −07 | −06 | 02 | 08 | −01 | 10* | 00 | 08 | −03 | −02 |
| Others | −10 | 01 | 07 | −15 | −09 | 08 | −02 | 17** | −01 | 00 |
| Children | −11 | −04 | 14 | −02 | −17 | 07 | −04 | 04 | 01 | 05 |
| Authority | −11 | 22* | 09 | 05 | −07 | 12* | 03 | 02 | 00 | 00 |
| Work | −10 | 24* | 07 | 02 | 07 | 11* | 02 | 06 | −05 | −05 |
| Reality | −07 | 14 | 01 | −11 | 08 | 19** | 01 | 11* | 04 | −03 |
| Parents | 06 | 03 | 03 | 05 | 00 | 09 | 04 | 12* | −14** | −04 |
| Hope | −14 | 13 | 12 | −07 | −14 | 05 | −05 | 16** | −02 | −02 |
| Total | −12 | 12 | 11 | −02 | −07 | 15** | 00 | 15** | −05 | −02 |
| Ratio | 03 | 00 | 05 | −03 | 16 | −07 | 01 | 03 | −05 | 03 |
| *R* | .209 | .397 | .171 | .244 | .293 | .209 | .100 | .237 | .185 | .102 |
| *F* | .34 | 1.40 | .23 | .47 | .70 | 1.68 | .37 | 2.19* | 1.30 | .39 |

Note.—Males $N = 86$; females $N = 379$.
* $p < .05$.
** $p < .01$.

Of the 180 correlations computed with CPI scales, 34 exceed the 5% level of confidence. On the other hand, none of the coefficients exceeds .30 in absolute magnitude. For SRI scores, 13 of the 100 coefficients exceed the 5% level, but none are larger than .24. There appear to be some minor relationships between these paper-and-pencil inventories and the behavioral evaluations by the pupils, but they are quite indistinct at best. In terms of multiple correlations, the results are equally unimpressive. For males, only Factor IV was reliably predicted by CPI scales, while none of the factors could be predicted by SRI scores. For females, Factors I, II, and III were reliably predicted by CPI, but only Factor III could be predicted by SRI.

The first POSR factor (Friendly, Cheerful, Admired) is related only to CPI Dominance scale in a negative direction for males. For females it is related positively to CPI Sociability and negatively to Psychological-mindedness. It is also related positively to SRI Total and Reality scales. These relatively strong correlations suggest that, at least for females, the first POSR factor does indeed reflect a comfortable, positive acceptance of life and an interest in other people on a social as opposed to an intellectual level. For males this factor may involve a quite different personality pattern, since 15 of the 18 scales of CPI and 9 of the 10 SRI scores show reversed signs on the correlation coefficients in comparing males with females. Although the evidence is only suggestive, this factor appears to reflect a generalized timid uncertainty among male student teachers.

The second POSR factor (Knowledgeable, Poised) was related to four scales reflecting a stable orientation toward work, among males. For females it was most strongly related to CPI Dominance, and also to scales reflecting a desire to achieve within a traditional authority structure.

The third POSR factor (Interesting, Preferred) is most strongly related to what might appear to be dependent intellectual attitudes for both sexes; the correlations were negative with Achievement via Independence, Psychological-mindedness, and Flexibility. For females only, this factor was also related positively to CPI Self-acceptance and SRI Total, Hope (for the future), and Others scales. Pupils appear to prefer teachers who are lively and interesting, and these people would seem to be characteristically nonrebellious and positively accepting of life as they find it.

The fourth POSR factor (Strict Control) yielded results suggesting that, when considered apart from the levels on the other POSR factors, this dimension reflects characteristics most psychologists would evaluate negatively. For males the negative correlations with CPI Achievement via Independence, Intellectual Efficiency, and Psychological-mindedness seem to reflect a syndrome of anti-intellectual attitudes. For females the negative correlation with SRI Parents, combined with negative correlations with CPI Good Impression and Femininity, may well indicate an etiological clue to a syndrome of "bitterness."

The implication here is that negative attitudes toward parents, combined with insensitivity and self-centeredness (suggested by low Femininity and Good impression scores), produce behavior in a teaching role which is seen by pupils as rather harshly overcontrolling. Further research is needed to confirm these relationships in other samples, however.

The fifth POSR factor (Democratic Procedure) was unrelated to any of the scales for males, but was positively related to CPI Responsibility and, understandably, to Flexibility for females. One might speculate that this factor reflects the attempts of some student teachers to conscientiously apply democratic principles in managing their classes.

Validation against supervisor ratings of effectiveness. In Table 7 are located the comparisons between supervisor ratings and POSR factor scores. Means on each POSR factor are included for each sex, at each of the three levels of teaching effectiveness, as determined by the

TABLE 7

ANALYSES OF POSR FACTOR SCORES AT THREE LEVELS OF TEACHING EFFECTIVENESS (E) AS DEFINED BY SUPERVISOR RATINGS

| POSR factor | High E *M* | Average E *M* | Low E *M* | *F* |
|---|---|---|---|---|
| | Male student teachers | | | |
| I. Friendly, Cheerful, Admired | 43.66 | 50.47 | 44.50 | 4.36* |
| II. Knowledgeable, Poised | 52.84 | 50.16 | 46.62 | 3.16* |
| III. Interesting, Preferred | 49.57 | 51.26 | 49.37 | .23 |
| IV. Strict Control | 55.65 | 47.90 | 45.05 | 6.69** |
| V. Democratic Procedure | 49.43 | 49.18 | 48.39 | .07 |
| *N* | 24 | 29 | 34 | |
| | Female student teachers | | | |
| I. Friendly, Cheerful, Admired | 53.48 | 51.48 | 47.27 | 13.68** |
| II. Knowledgeable, Poised | 52.77 | 49.57 | 47.74 | 8.43** |
| III. Interesting, Preferred | 50.65 | 49.57 | 48.71 | 1.30 |
| IV. Strict Control | 51.56 | 50.05 | 48.35 | 3.66* |
| V. Democratic Procedure | 51.15 | 51.43 | 49.41 | 1.78 |
| *N* | 127 | 151 | 132 | |

* $p < .05$.
** $p < .01$.

ratings made by supervising professors.

For both male and female student teachers, supervisor ratings of effectiveness are a function of the same three factors: I. Friendly, Cheerful, Admired; II. Knowledgeable, Poised; and IV. Strict Control. Except for the males on Factor I, all of these relationships appear to be linear and positive.

In contrast to the females for whom greater friendliness is related to a higher effectiveness rating, the male student teachers who receive the highest scores on friendliness are those rated as mediocre in effectiveness. It will be recalled that this first factor also yielded a series of negative correlations with various attitude scales for males, but positive correlations for females. Evidently a high rating on Factor I implies quite a different pattern of other qualities for the two sexes. For females it is associated with social confidence and effective teaching, while for males it is found among somewhat timid individuals of mediocre effectiveness.

## DISCUSSION

The cliché that you get out of a factor analysis somewhat less than you put into it probably holds in the case of the present investigation. Nevertheless, the factor dimensions defined here do have certain properties which give them considerable promise for further research with teacher candidates, either as criteria or as predictors. Their internal consistency is, in one sense, guaranteed by the manner in which they were defined, while their invariance over subject populations and pupil subclasses has been adequately demonstrated.

As pointed out by Stern (1963), strikingly similar factors have emerged from a number of independent analyses of student ratings. He concluded that "the verbalized teacher image is a widely shared and extremely stable stereotype." Empathy (friendly, democratic behavior) and competence (systematic, organized behavior) were mentioned as the essential components of this idealized image. In the present analysis five independent aspects of teacher behavior were identified; Factors I and V might fall under the "empathy" heading, while Factors II and IV would appear to be related to "competence." Factor III (Interesting, Preferred), however, probably corresponds more clearly with the pupil's notion of "effective teaching" than any of the other factors.

As was described earlier in the present report, there appears to be a close correspondence between each of Ryans' three patterns and one or another of the POSR factors. This would seem to confirm the hypothesis that pupils can provide at least as much information regarding teacher characteristics as can expert adult judges on the basis of 1 or 2 hours of observation. There are a number of indications in the present study, however, of the need for a thorough study of the relationships between the value systems of adolescent judges and adult experts regarding teacher characteristics.

The most striking result of the comparisons of supervisor rated effectiveness and POSR factors was the total lack of relationship between supervisor evaluations and Factor III: Interesting, Preferred. It seems quite apparent from the item content shown in Table 1 that this factor reflects the pupils' ideas of an effective teacher. The supervisor's ratings, on the other hand, are quite clearly a function of the student-teacher's friendliness (I), poise, (II), and control of the classroom (IV). It is also interesting to note the lack of correspondence between Factor V (Democratic Procedure) and the effectiveness ratings. Inspection of the means for the three groups suggests the hypothesis that autocratic behavior may be related to poor teaching, but lends little support to the idea that democratic methods are a necessary part of effective teaching behavior—at least as supervising professors see it. This hypothesis deserves more definitive testing in future research.

The reasons for the low predictive validity of the attitude inventories used in this study also deserve concern in further research. Self-report questionnaires simply may not provide a sufficient degree of differentiation to yield anything but slight indications of relevant personality traits, due to restriction of range of these attributes in the present samples. It should be remembered that all of the subjects in the present study were college seniors who eventually graduated and were certified as competent teachers.

A number of other questions remain to be answered regarding possibly important sources of variation in POSR factor scores. Among these potential influences are the interactions of pupil and teacher sex (Velman & Peck, in press), the social class level of the school in which the teaching was done, the grade level of the pupils, and particularly the interaction of teacher and pupil personality characteristics (Heil, Powell, & Feifer, 1960). There is no reason why factor scores need be computed only from the item means for teachers. Differences in viewpoint among the members of a given class may also be revealed by separate computation of scores on the five dimensions, from each pupil's protocol.

## REFERENCES

AHMAVAARA, Y. Transformation analysis of factorial data. *Ann. Acad. Sci. fenn.*, 1954, 88, No. 2.

BARR, A. S., WORCESTER, D. A., ABELL, A., BEECHER, C., JENSEN, L. E., PERONTO, A. L., RINGNESS, T. A., & SCHMID, J., JR. *Wisconsin studies of the measurement and prediction of teacher effectiveness.* Madison: Dembar, 1961.

BOWN, O. H. The development of a self-report inventory and its function in a mental health assessment battery. *Amer. Psychologist*, 1961, 16, 402.

DALTON, E. L. *What makes effective teachers for young adolescents?* Nashville: Department of Education, George Peabody College for Teachers, 1962.

GOUGH, H. G. *Manual for the California Psychological Inventory.* Palo Alto, California: Consulting Psychologists Press, 1957.

HEIL, L. M., POWELL, MARION., & FEIFER, I. Characteristics of teacher behavior and competency related to the achievement of different kinds of children in several elementary grades. May 1960, Brooklyn College, Contract No. SAE 7285, United States Office of Education Cooperative Research.

MCCLAIN, E. The relationship between student teachers' self-reported perceptions and pupil evaluations. Unpublished doctoral dissertation, University of Texas, 1961.

MCCLAIN, E. & BOWN, O. H. Personality characteristics of student teachers whose self-descriptions differ markedly from descriptions by their pupils. *Amer. Psychologist*, 1961, 16, 403.

RYANS, D. G. *Characteristics of teachers.* Washington, D. C.: American Council on Education, 1960.

SARASON, S., DAVIDSON, K., & BLATT, B. The *preparation of teachers: An unstudied problem in education.* New York: Wiley, 1962.

STERN, G. G. Measuring noncognitive variables in research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching.* Chicago: Rand McNally, 1963.

VELDMAN, D. J., & PECK, R. F. The influence of teacher and pupil sex on pupil evaluations of student teachers. *J. teacher Educ.*, in press.

(Received May 15, 1963)

*Journal of Educational Psychology*
1963, Vol. 54, No. 6, 356–362

# LEARNING UNDER CONDITIONS OF DIRECT AND VICARIOUS REINFORCEMENT[1]

R. KEITH VAN WAGENEN AND ROBERT M. W. TRAVERS
*Arizona State University* *University of Utah*

This study is concerned with the effect of various reinforcing contingencies on the learning of the English equivalents for 60 German words. This task was administered in a situation involving groups of 8 Ss and a "teacher" E. In addition, 2 learning conditions, with Ss learning in isolation, involved the use of teaching machines. In the group situation 3 different relationships to the reinforcement were introduced. Ss who made no overt response, but simply observed, learned reliably less than Ss in the same group who learned with overt responses followed by a reinforcing event.

The classroom is rarely a suitable place for conducting studies which may contribute to an understanding of the learning process, for the events which occur within it include many uncontrolled conditions which may mask the particular effects in which the experimenter is interested. On the other hand, research conducted within the laboratory is undertaken under conditions so different from those of the classroom that one is hardly justified in applying knowledge derived from such experimentation to the solution of problems of classroom learning. Laws of learning discovered under typical laboratory conditions have boundary conditions, and one cannot assume that learning in the classroom falls within these boundaries unless this has been demonstrated. What appears to be needed for the solution of educational problems is research conducted in situations which retain all the desirable features of laboratory experimentation, but which also simulate certain characteristics of the classroom. The results of such studies can be generalized to classroom settings with much less fear that the generalization is unwarranted.

Laboratory research involving situations related to classroom learning would appear to be needed particularly at this time in the area of reinforcement learning, for experimental psychologists have shown a proneness to generalize research results derived from the study of infrahuman subjects in highly controlled laboratory situations to the solution of problems of teaching children in schoolrooms.

The central part of the study reported here is an attempt to investigate the effect of certain classes of reinforcers in a situation in which a high degree of control can be exercised over the presentation of stimuli related to learning, the response conditions, and the reinforcing contingencies. The study is focused on the phenomenon which has been referred to by Osgood (1956) and deCharms and Rosenbaum (1960) as vicarious learning. *Vicarious learning* is defined as learning occurring in an "observer" organism which has the opportunity of watching another member of the same species. Proponents of reinforcement learning theory have generally taken the position that learning will be most effective if the learner makes the full response to be learned and is reinforced for it and that such learning will be greatly superior to that of an observer who makes no overt response and for whom there can be no clearly identifiable reinforcement contingent on the response. A number of investigations with infrahuman subjects, such as those by Miller and Dollard (1941), Angermeier, Schaul, and James (1959), Darby and Riopelle (1959), and Hayes and Hayes (1952), have demonstrated the existence of vicarious learning at various phylogenetic levels. Other studies such as those of Kounin and Gump (1958), Lewis and Duncan (1958), Rosenbaum (1962), and Rosenblith (1959) have examined the vicarious learning in humans.

---

[1] This research was supported by Office of Education Research Contract OE No. 2-10-010 to R. M. W. Travers.

The typical classroom recitation situation is one in which the teacher anticipates that vicarious learning will take place. Recitation involves the following pattern of pupil-teacher interaction:

1. The teacher addresses a question to Pupil X.
2. The pupil gives an answer.
3. The teacher provides a reinforcement.

The procedure generally involves the assumption that, as a result of this interaction, other pupils not included in it, but proximate to the respondant, will also learn and to about the same extent as Pupil X. The validity of the latter assumption is examined in this research.

In addition, data are provided comparing the performance of the directly reinforced pupil with the performance of pupils learning under another condition which has been widely advocated as being particularly effective by reinforcement theorists. This is the learning condition provided by a teaching machine which is claimed to provide conditions for learning superior to those provided by the classroom recitation procedure in that the subject responds fully and overtly to each situation presented and reinforcing events are provided contingent on each response.

However, since the teaching machine provides reinforcement through a mechanical arrangement, another question raised is whether human reinforcing agents have properties which mechanical reinforcing agents do not have. Hence, the study makes some provision for studying this question.

Finally, there is the possibility to be explored that an occasional pupil-teacher interaction may maintain the overall learning of the pupil at a higher level than it would be without such an intermittent exchange.

## METHOD

### *Task*

The first step which had to be undertaken in the development of this study was the construction of a learning task which (*a*) could be administered in a learning situation conforming to the response-reinforcement learning model, (*b*) was similar to learning tasks commonly observed in classrooms, (*c*) was such that it could be administered under several different learning and reinforcement conditions, and (*d*) would permit a reliable measure of learning. The task selected was that of learning and meaning of 60 German words.

The learning task consisted of a German stimulus word which in the experiment was to be correctly associated with one of two English noun responses. German was selected because it was assumed that there would be few children with any familiarity with this language, and in addition, that interest in learning foreign words would be high.

The general format of the task required a subject to guess which one of two English words meant the same as a given German word. Reinforcement followed the response. These pairs of English nouns (one of which was the equivalent of the German stimulus) had to be selected so that one could assume that they had equal association to the German stimulus. In providing a pool of such words all nouns of three to eight letters in length (not having some obvious association with the German word) were taken from the Thorndike-Lorge (1944, Appendix, Part V) list of Original Thorndike Words. The pairs were selected separately for the first 500 most frequently occurring and for the second 500 most frequently occurring words. An additional 7 were taken from the list of 110 appended to the original 1,000, which yielded a total of 314 words from which the task could be assembled.

On the basis of an argument advanced by Underwood and Schulz (1960), a second step in providing for equal association value was the assignment of a numerical frequency value for the first bigram of each selected word, i.e., (AP)PLE. In the pairing of response choices the bigrams of the two words had to have approximately comparable frequencies (as determined by the compilers, Underwood and Schultz).

In addition to the procedures already outlined, each English noun was checked for level of readability to insure that words used were well within the reading capacity of the participating subjects. Durrell (1956), in his Appendix A of *Improving Reading Instruction,* lists 754 words compiled at the primary grade level as vocabulary for older children in remedial reading programs. Each word has been assigned a value ranging from 1 to 7 which indicates frequency of occurrence in the literature of small children. It was found that 69% of the response nouns in the learning task is listed among Durrell's words. As far as was possible, pairs of response choices were matched with respect to Durrell's readability values. Those nouns already selected for the task, which were not part of the Durrell list, were checked against another Durrell list of vocabulary to be learned in the fourth grade. While not many of the learning task nouns were found there, it was obvious that Durrell's fourth grade list was at a higher level of difficulty. It was concluded that the words selected would be well within the range of reading ability for the proposed sample of subjects.

The task was prepared in two different forms. In one form, suitable for administration in a group situation, each German word and the accompanying English words were printed in large block letters on sheets of card 11 × 14 inches. In another form the materials were typed on paper suitable for insertion into a teaching machine (Koncept-o-graph No. 7).

The acquisition trials were conducted on each of 3 successive days: Monday, Tuesday, and Wednesday. On each of these days a set of 20 words was presented with four trials on each. The total of 80 presentations per day required approximately 40 minutes to complete. The Criterion Measure was not administered until Friday. The test called for recognition of the English equivalent among four choices. The three distracters included the word used as an incorrect choice in the learning task and also two English words derived from other task items.

## *Task Administration*

The task was presented with two major variations, an adaptation for learning in a group, and one for learning in isolation. The first of these variations involved presentation of the task to a classroom group consisting of eight "pupil"-subjects seated in a familiar classroom row and facing the experimenter who functioned as the teacher. The "teacher"-experimenter stood approximately 7 feet from the center of the row, from which position she interacted with the odd numbered subjects in the row but not with those evenly numbered, except on a single practice trial which required participation by all subjects. Instructions were given concerning the general nature of the task and how one of the two English words which meant the same as the German word was to be selected. In the administration of the actual learning task the teacher-experimenter held up a card on which was printed a German word. [illegible]n made a statement such as the following: "John, this word is BAUM." The pupil-subject, John, selected one of two English words, alternatives as the equivalent of Baum, which appeared on the card held by the experimenter. If the subject said "TREE," the correct choice, the experimenter responded, "right" or "that is correct." When he was wrong the experimenter said nothing. The remainder of the group listened to the exchange. During the learning trials of that day John would have been confronted with this same item a total of four times. In all, he was presented with five such items, making a total of 20 presentations to which he responded on a single day. The nature of the task was such that only one subject at a time could be directly involved in responding and receiving feedback. This meant that three of the four subjects who vocally interacted with the experimenter were indirectly, or vicariously, involved with an item which went directly to the fourth. The experimenter who presented the learning task to the group was a young woman who had several years of teaching experience.

The task and the conditions under which it was administered permitted the study of learning under the following conditions:

*Condition A: Partial Verbal Interaction between Experimenter and Subject.* The experimenter interacted with only the odd numbered subjects, but, since these four subjects responded in turn, interaction was on only one fourth of the vocabulary items. These same subjects, and odd numbered ones, also had the opportunity of learning the remaining items of the task (to which they did not respond directly) by observing the interactions with the experimenter of the remaining odd numbered subjects. Thus, they could learn one fourth of the items in the task through oral responding and reinforcement, while having the opportunity to learn the remaining three fourths of the task through what is termed here vicarious learning.

*Condition B: No Verbal Interaction.* Within the simulated classroom group a third learning condition existed. The even numbered subjects had no interaction with the experimenter during the entire learning task. Whatever learning took place in them fell in the category of vicarious learning.

The task was also arranged so that it could be presented to subjects who were to learn in isolation through a Koncept-o-graph teaching machine which utilized sheets of paper (8½ × 11 inches) on which the items were typed. Two different conditions of learning were introduced [illegible] are labeled here C and D.

*Condition C: Isolation with Machine Feedback.* The subject was seated in front of a teaching machine with only himself and the experimenter in the room. The machine presented a German word together with the two English words. The subject made his choice, but neither wrote it down nor did he vocalize it. Following each item, the machine provided knowledge of results.

*Condition D: Isolation with Aural As Well As Machine Feedback.* This condition was identical with the previous one except that the subject was asked to say his response while the experimenter was seated beside him. If the response was correct, the experimenter said, "right," "that is correct," or "correct." If the response was incorrect, the experimenter said nothing. In addition, the machine displayed the correct response.

### *Sample*

Participating subjects were 180 elementary school children in the fourth, fifth, and sixth grades (91 boys and 89 girls). All subjects were naive as to the purpose of the study and had no acquaintance with the German language. Participants were assigned to particular treatments on the basis of a reading score (Metropolitan, Iowa, or Science Research Associates, as available in the school). A reading score was selected for equating groups because: (*a*) performance on the task required some reading competence, and (*b*) reading ability and verbal learning measures have been demonstrated to be highly correlated (such tasks do not place high premium on intelligence—Wechsler, 1944).

TABLE 1

DISTRIBUTION OF SUBJECTS BY TREATMENT

| Treatment | *N* |
|---|---|
| Subjects in simulated classroom | |
| A. Subjects partially interacting with teacher-experimenter | 48 |
| B. Subjects not interacting with teacher-experimenter | 48 |
| Subjects learning in isolation | |
| C. With feedback from the teaching machine only | 44 |
| D. With feedback from both teaching machine and experimenter | 40 |

Table 1 presents the distribution of subjects according to learning conditions. Altogether 12 simulated classroom groups of eight pupils in each were involved. Three elementary schools in Salt Lake City and one in a suburb provided subjects for the experiment.

## RESULTS

The data consisted of scores derived from a multiple-choice test of German vocabulary administered to those who had been exposed to the various conditions. This test was administered on the Friday of the same week when the three learning sessions had taken place on Monday, Tuesday, and Wednesday. The test could be scored as a whole, or three separate scores could be derived measuring separately the learning accomplished during each one of the learning sessions. As a first step in the study of the data, an analysis of variance was undertaken involving the four conditions of learning, A, B, C, and D, as one set of comparisons and the three scores pertaining to the three learning sessions for another. The first 40 subjects in each treatment were used for this analysis. Table 2 summarizes the analysis of

TABLE 2

ANALYSIS OF VARIANCE FOR LEARNING CONDITIONS AND DAYS OF THE LEARNING TRIALS

| Source | df | MS | F | Required .05 | Required .01 |
|---|---|---|---|---|---|
| Between days (lists) | 2 | 777.0 | 82.7 | 2.99 | 4.60 |
| Between learning conditions | 3 | 30.7 | 3.3 | 2.68 | 3.95 |
| Interaction | 6 | 33.0 | 3.5 | 2.17 | 2.96 |
| Within | 468 | 9.4 | | | |
| Total | 479 | | | | |

variance. It is obvious that the amount of learning manifested at the time of the criterion test depends upon the particular day in the sequence when the learning originally occurred. It is also evident that the differences between learning conditions surpassed the 5% level of significance. The interaction of learning condition and days is significant beyond the 1% level. Brief consideration must be given to each one of these significant differences.

*Differences between Days*

Retention scores on the Friday test showed a very marked decline from the first day through the second to the third day of learning. The task learned first was the one best remembered, and this applied to all learning conditions.

*Differences between Learning Conditions*

That significant differences exist at least at the 5% level of significance is clear from the analysis of variance. Table 3 provides information concerning the nature of this effect. Conditions A and C indicate closely similar mean performances which are superior to Conditions B and D which are very near to one another. Table 4 provides a Scheffe test which indicates the level of significance for individual comparisons.

Some results of additional interest are provided by a further breakdown of Condition A into the two component treatments which it involves. Under Condition A the experimenter interacted with the learner during the presentation of 5 of the vocabulary items per day of trials, and did not interact on the other 15 items. On those items for which

TABLE 3

MEAN TEST PERFORMANCE ON FRIDAY FOR EACH DAY'S TASK

| Task | Simulated classroom with teacher interaction | Simulated classroom no teacher interaction (vicarious learning) | Isolated machine feedback only | Isolated machine feedback plus teacher feedback |
|---|---|---|---|---|
| Monday (20 items) | 12.80 | 11.85 | 11.73 | 10.03 |
| Tuesday (20 items) | 8.40 | 8.03 | 9.18 | 9.43 |
| Wednesday (20 items) | 7.40 | 6.43 | 8.33 | 6.90 |
| Total for entire task | 28.59 | 26.31 | 29.22 | 26.36 |

TABLE 4

Scheffe Test of Multiple Comparisons for Learning Conditions Table 1

| Source | F | Required .05 | Required .01 |
|---|---|---|---|
| A versus B<br>Direct ¼+¾ versus vicarious 4/4 | 11.28 | 8.01 | 11.73 |
| A versus D<br>Direct ¼+¾ versus isolated—aural feedback | 10.74 | 8.01 | 11.73 |
| B versus C<br>Vicarious 4/4 versus isolated | 18.19 | 8.01 | 11.73 |
| C versus D<br>Isolated versus isolated—aural feedback | 17.55 | 8.01 | 11.73 |

there was direct interaction, the probability of a correct response on the Friday test was .54, while on the items vicariously experienced (no verbal interaction) the probability was .46. These two treatments within Condition A were significantly different in learning at the .002 level (*t* test with correlated data). It is also of interest to note that the items on which the subjects interacted with the experimenter produced more learning than that produced by any of the learning conditions studied. This fact is made apparent in Figure 1 (top) where Condition A is shown in comparison with Condition C. This graph suggests that there is least decline over the 3 days when the learning condition fits the response-reinforcement learning paradigm, and most when learning is a relatively passive process.

A number of other comparisons within the data are worth noting. Under Condition A three fourths of the items were vicariously experienced (did not involve an interaction with the experimenter). It is of considerable interest to compare the learning of these items within Condition A with the learning which occurs in Condition B and which involved *only* vicarious learning. This comparison is shown in Figure 1 (bottom) which presents the amount of vicarious learning in terms of the probability of answering the items correctly in the test on Friday. Apparently, the advantage of Condition A is restricted to those items on which there was a verbal exchange between the experimenter and the subject, since the two vicarious conditions appear not to differ with respect to learning.

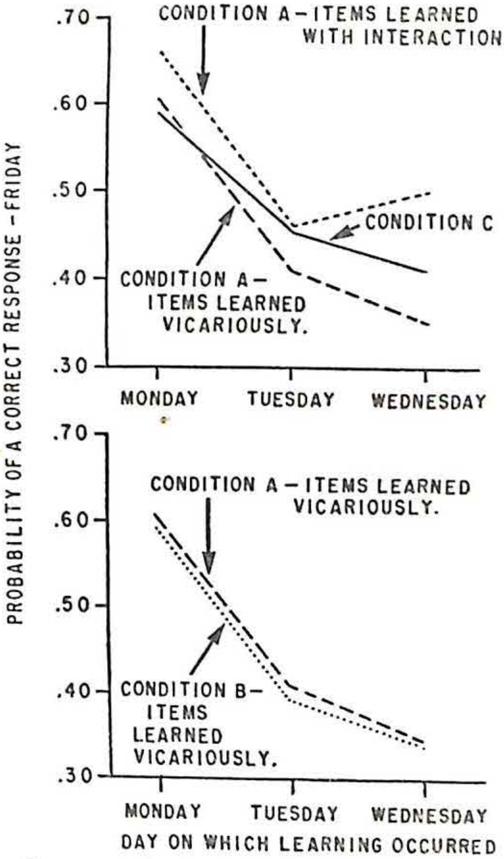


Fig. 1. Criterion test performance (measured Friday for each week of experimenting), showing separate retention of learning for Monday's, Tuesday's and Wednesday's tasks.

### *Interaction between Learning Conditions and Days*

This interaction is best understood by reference to the means presented in

Table 3. The means by days show that there is a more rapid decline in performance in the simulated classroom situation (Conditions A and B) than in the teaching machine situation (Conditions C and D). A final point to note from Table 3 is that the vicarious learning situation appears to show a greater decline over the 3 days than does any other learning condition. The data are not sufficient to make any great issue of this point, but they indicate a line for further research.

## DISCUSSION

The data reported here make it possible to compare the effectiveness of learning under various conditions including those commonly observed in classrooms and also under conditions which approximate closely those of the reinforcement paradigm. While the advocates of the application of the reinforcement paradigm to the planning of school learning have claimed that such application would produce dramatic increases in the learning of pupils, the results of the study do not support this position. Those subjects who interacted with the experimenter and whose learning behavior followed the familiar reinforcement paradigm obtained a score of 54% on Friday on those items on which they had direct interaction, while the vicarious subjects who never interacted with the experimenter during the administration of the task achieved a score of 44.3%. The two conditions probably represent the best and worst conditions of learning commonly observed in schools. However, the fact that Condition B, learning entirely by vicarious experience, showed such a marked decline in performance suggests that for this group the situation becomes dull and progressively less capable of commanding task orientation. While the decline in performance in all groups from Monday through Tuesday to Wednesday is marked, and may well be due partly to a proactive inhibition effect, the particularly striking decline of the vicarious group suggests that an attention factor is involved and may be partially responsible for the effect. It would be of some interest to prolong such a learning series over a much longer period than 3 days to determine whether the deterioration of the vicarious learning group would become progressively more marked. One suspects that it would.

## REFERENCES

ANGERMEIER, W. F., SCHAUL, L. T., & JAMES, W. T. Social conditioning in rats. *J. comp. physiol. Psychol.*, 1959, **52**, 370–372.

DARBY, C. L., & RIOPELLE, A. J. Observational learning in the rhesus monkey. *J. comp. physiol. Psychol.*, 1959, **52**, 94–98.

DECHARMS, R., & ROSENBAUM, M. E. The problem of vicarious experiments. In D. Willner (Ed.), *Decisions, values and groups.* New York: Pergamon Press, 1960.

DURRELL, D. D. *Improving reading instruction.* New York: World Book, 1956.

HAYES, K. J., & HAYES, C. Imitation in a home raised chimpanzee. *J. comp. physiol. Psychol.*, 1952, **45**, 450–459.

KOUNIN, J. S., & GUMP, P. V. The ripple effect in discipline. *Elem. Sch. J.*, 1958, **59**, 158–162.

LEWIS, D. J., & DUNCAN, C. P. Vicarious experience and partial reinforcement. *J. abn. norm. soc. Psychol.*, 1958, **57**, 321–326.

MILLER, N. E., & DOLLARD, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

OSGOOD, C. E., Behavior theory and the social sciences. *Behav. Sci.*, 1956, **1**, 167–185.

ROSENBAUM, M. E. Effect of direct and vicarious verbalization on retention. *Child Develpm.*, 1962, **33**, 103–110.

ROSENBLITH, JUDY F. Learning by imitation in kindergarten children. *Child Develpm.*, 1959, **30**, 69–80.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Columbia Univer. Press, 1944.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

WECHSLER, D. *The measurement of intelligence.* Baltimore: Williams & Wilkins, 1944.

(Received June 17, 1963)